A. Mishchenko and A. Fomenko

A Course of Differential Geometry and Topology

A Course of Differential Geometry and Topology

А. С. Мищенко, А. Т. Фоменко

КУРС ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ И ТОПОЛОГИИ

ИЗДАТЕЛЬСТВО МОСКОВСКОГО УНИВЕРСИТЕТА

A. Mishchenko and A. Fomenko

A Course of Differential Geometry and Topology

Mir Publishers Moscow

Translated from Russian
by Anatoly Talashev

First published 1988
Revised from the 1980 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-03-000220-0

Contents

Preface to the English Edition

The English edition has been thoroughly revised in line with comments and suggestions made by our readers, and the mistakes and misprints that were detected have been corrected. This is essentially a textbook for a modern course on differential geometry and topology, which is much wider than the traditional courses on classical differential geometry, and it covers many branches of mathematics a knowledge of which has now become essential for a modern mathematical education. We hope that a reader who has mastered this material will be able to do independent research both in geometry and in other related fields. To gain a deeper understanding of the material of this book, we recommend the reader should solve the questions in A.S. Mishchenko, Yu.P. Solovyev, and A.T. Fomenko, *Problems in Differential Geometry and Topology* (Mir Publishers, Moscow, 1985) which was specially compiled to accompany this course.

Preface to the Russian Edition

Differential geometry and topology constitute an important branch of mathematics. The rapid progress in theoretical physics and mechanics during the 20th century has shown that geometric concepts lie at the heart of certain fundamental theories. Suffice it to point out such fields as relativity theory, the mechanics of continuous media, and electrodynamics, where the mathematical apparatus rests upon geometric methods.

The present course deals with the fundamentals of differential geometry and topology whose present state is the culmination of contributions from generations of mathematicians. It is rather difficult to survey the history of geometry because we now treat most geometry problems quite differently from the way mathematicians did in the past.

Almost all the methods of geometry (and geometry is not the only such field) have been reworked. From the modern point of view, geometry and topology are based on several fundamental ideas, which may be outlined as follows.

Set-theoretic (or general) topology is the branch of geometry that has "absorbed" the once general methods for studying convergence and continuity. One important idea in geometry—the use of curvilinear coordinate systems—eventually led to tensor analysis and the theory of invariants. Whereas mathematical analysis and the theory of differential equations deal primarily with the "local" properties of a function (only infinitely adjacent points are considered), geometry studies the "global" properties of functions (that is, their properties are analysed at finitely spaced points). This intuitive idea of the study of the "global" properties of geometric objects has given rise to the fundamental concept of manifold as a generalization of the concept of domain in Euclidean space.

Later, the "marriage" of these ideas in various particular applications became equally important in geometry: the geometric interpretation of integration, cohomology theory as a formal geometric representation of an intuitive transition from the "local" to a "global" study of a space, geometric interpretation of a linear approxima-

tion of non-linear functions and mappings, the principle of general position in geometry.

All these geometric ideas were invented to facilitate the analysis of particular problems, and they have furthered both the solution of these problems and the development of geometric concepts and methods. It is sometimes rather difficult to distinguish between the ideas that are now concerned with differential geometry and topology and those that have developed into separate directions of analytic geometry and linear algebra, or even belong to some other branches of mathematics.

Ancient geometry culminated in Euclid's *Elements* and had little influence on the development of differential geometry. Only one problem, which for many centuries seemed to be the most important, found a simple and natural solution within the framework of differential geometry. We mean the proof of Euclid's fifth axiom. The axiom was so obscure and so intimately related to the other axioms that for many centuries mathematicians vainly attempted to prove it by proceeding from the others. In all probability, Leo Gersonide (1288-1344) was the first mathematician in Europe who tried to prove the parallel axiom. The proof of the fifth axiom became the focus of keen attention by famous mathematicians such as Ch. Clavius (1574), P. Cataldi (1603), G.A. Borelli (1658), J. Vitali (1680), J. Wallis (1663), G.G. Saccheri (1733), J.H. Lambert (1766), A.M. Legendre (1800), F.K. Schweikart (1818), F. Taurinus (1825), and even C.F. Gauss. Although their attempts were unsuccessful, they were of tremendous significance, for they laid the foundations for a new, non-Euclidean geometry. This geometry was based on a rejection of the fifth axiom and was invented by N.I. Lobachevsky (1792-1856), the great Russian mathematician. In 1826 he delivered his first communication on non-Euclidean geometry, and this opened up a new era in the development of geometry. In 1832, a similar paper of J. Bolyai was published. The same ideas were put forward by C.F. Gauss, who did not, however, publish his work.

Whereas Lobachevsky arrived at his non-Euclidean geometry by the axiomatic method, further new ideas in geometry were developed concurrently with other fundamental principles of defining geometry objects via a coordinate system. Coordinates were employed even in ancient geography and astronomy. In Ptolemy's *Geography*, we come across latitude and longitude as numerical coordinates. In Europe, N. Oresme (1323-1382) came very close to the idea of coordinates in the graphic representation of a function, but he never mentioned the term "coordinates". The discoverers of the coordinate method and analytic geometry were R. Descartes (1596-1650) and P Fermat (1601-1665). The consistent use of curvilinear coordinates and Lobachevsky's ideas led to the rapid development of differential geometry. C.F.B. Riemann (1826-1866) took a novel step when

he pioneered the study of arbitrary, so-called Riemannian spaces, which were found to have wide applications in mechanics, relativity theory, and other fields. Riemann's studies stimulated vigorous growth of vast new branches of geometry: vector and tensor calculus and Riemannian geometry.

At the same time, the foundations of topology were laid as part of geometry dealing with continuity properties. Topology, of course, owes its origins to the infinitesimal calculus, although the first purely topological papers were written by L. Euler (1707-1783) and C. Jordan (1838-1922). At the beginning of the 20th century, the fundamental principles of topology were elaborated by M. Frechet (1878-1973), F. Hausdorff (1868-1942), H.L. Lebesgue (1875-1941), and L.E.J. Brouwer (1881-1966). J.H. Poincaré (1854-1912) contributed a lot to progress in topology and its applications. Nowadays, topology is a rapidly growing branch of mathematics.

L. Euler and G. Monge (1746-1818) initiated an independent study within differential geometry: surface theory. The union of the concepts of non-linear coordinates, vector and tensor calculus, and surface theory, as well as the fruitfulness of geometric ideas in the natural sciences led to the most fundamental notion in geometry, the manifold. Poincaré was apparently the first to assign a clear meaning to the notion. The concept of manifold has now taken such firm roots in mathematical thinking that mathematicians are no longer deliberately developing manifold theory.

Finally, we should mention one more thing that has played an important role in the evolution and understanding of geometry and the key problems of modern theoretical physics. It is the use of group methods. These are based on the invariance of geometric objects with respect to a space symmetry group. Group methods emerged in geometry from the study of projective geometry and they are associated with the names of Poncelet, Staudt, Möbius, Plücker, and Cayley. In 1872, F. Klein (1849-1925) published a famous paper, later called "The Erlangen Program", in which he summarized advances in geometry and rigorously formulated the group principles for developing geometry. Group methods have strongly influenced differential geometry, topology, and the various applications of geometry in the other branches of mathematics and the natural sciences. It should be borne in mind however that the group technique is not the only geometric method.

Many mathematicians have contributed to differential geometry and topology. We believe it is worth noting that Russian and Soviet scholars have made a strong impact on this progress.

Chapter 1

Introduction to Differential Geometry

1.1. CURVILINEAR COORDINATE SYSTEMS. SIMPLE EXAMPLES

1.1.1. BACKGROUND

Let us consider an n-dimensional Euclidean space which is usually denoted by $\mathbf{R}^n$. We assume that this space is provided with Cartesian coordinates $x^1, \ldots, x^n$ which permit a unique determination of the position of any point in $\mathbf{R}^n$ by associating with it a set of real numbers, the coordinates relative to a fixed orthogonal basis formed by mutually orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. The idea of describing a point in an Euclidean space by a set of real numbers (which may also be considered as the coordinates of the radius vector emerging from the origin to the point) underlies analytic geometry which solves various geometric problems by purely algebraic methods. This important approach was first introduced (explicitly) into mathematics by Descartes in whose honour we now say "Cartesian coordinates". Algebraization of geometry has played a key role in the development not only of geometry as such but also of mathematics as a whole. We shall not concentrate on the problems which are easily and elegantly solved by algebraic-analytic methods (viz., classification of second-order surfaces in a three-dimensional space) and refer the reader to numerous courses of algebra and analytic geometry. Let us only recall that Cartesian coordinates in $\mathbf{R}^n$ are closely related to the concept of the Euclidean scalar product, a bilinear form which associates with each pair of vectors $\xi, \eta \in \mathbf{R}^n$ a real number usually denoted by $\langle \xi, \eta \rangle$; this operation is symmetric and linear in each argument, and the form itself is positive definite. In a Cartesian coordinate system we have

$$\langle \xi, \eta \rangle = \xi^1\eta^1 + \ldots + \xi^n\eta^n, \text{ where } \xi = (\xi^1, \ldots, \xi^n), \quad \eta = (\eta^1, \ldots, \eta^n).$$

Simple examples however show that Cartesian coordinates are not always the most convenient ones to solve analytically many particular problems. We shall demonstrate this by writing the equations of curves on a plane in Cartesian coordinates x, y. Of course, for rather simple curves, viz., a circle or ellipse, the analytic expressions in Cartesian coordinates are simple. Indeed, the equation of a circle of radius R with centre at point A is $(x - A^1)^2 + (y - A^2)^2 = R^2$, where $A = (A^1, A^2)$.

The equation of an ellipse is also simple: $(x - A^1)^2/a^2 + (y - A^2)^2/b^2 = R^2$, where a and b are the principal semi-axes (Fig. 1.1).

However, in various applications and concrete mechanical and physical problems we often deal with smooth curves (say, trajectories of the motion of a particle in a force field) whose equations

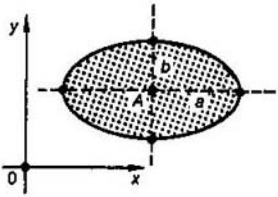

Figure 1.1

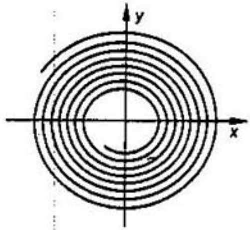

Figure 1.2

in Cartesian coordinates are rather cumbersome. For example, the equation $\sqrt{x^2 + y^2} - e^{\lambda\left(\tan^{-1}\frac{y}{x}\right)} = 0$ defines a spiral in Cartesian coordinates (Fig. 1.2). Although this equation is rather simple, it could be written in a simpler form in the so-called polar coordinates (r, φ) related to the Cartesian coordinates x, y by $x = r\cos\varphi$, $y = r\sin\varphi$ (Fig. 1.3). In polar coordinates the equation of a spiral be-

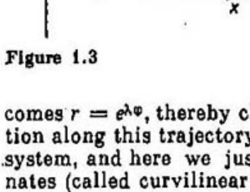

Figure 1.3

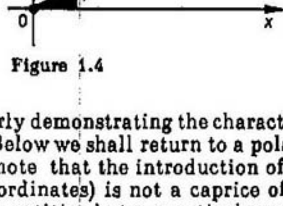

Figure 1.4

comes $r = e^{\lambda\varphi}$, thereby clearly demonstrating the character of the motion along this trajectory. Below we shall return to a polar coordinate system, and here we just note that the introduction of such coordinates (called curvilinear coordinates) is not a caprice of mathematicians trying to devise new entities but a practical necessity, sometimes quite useful in particular calculations. In this connection, we shall briefly discuss a problem where polar coordinates appear to be quite useful.

Let us consider the motion of a particle on a plane in a central force field. Suppose the centre is at the point O and (r, φ) are polar coordinates on the plane. Let $\mathbf{r}$ be the radius vector of the moving particle (this vector originates from O), r be its length, and t be time (the motion parameter); then the coordinates r and φ are functions of time. Consider at a point $\mathbf{r}(t)$ with polar coordinates $r = |\mathbf{r}|$, φ two orthogonal unit vectors: a vector $\mathbf{e}_r$ along the radius vector of the particle (note that the relation $\mathbf{r} = r \cdot \mathbf{e}_r$ holds in this case) and a vector $\mathbf{e}_\varphi$ orthogonal to $\mathbf{e}_r$ and so directed that the polar angle φ increases (Fig. 1.4). Differentiation of a radius vector $\mathbf{r}(t)$ with respect to time will be denoted by a dot. As is known from mechanics, the motion of a particle (its mass is taken equal to 1, for simplicity) in a central force field on a plane is described by the following differential equation: $\ddot{\mathbf{r}} = f(r)\,\mathbf{e}_r$, where f is a smooth function of a single argument r. Incidentally, here is a useful exercise for the reader: write this differential equation in Cartesian coordinates on a plane.

The motion of a particle can be described by two functions: $r = r(t)$ and $\varphi = \varphi(t)$, that is, in a polar coordinate system. It is a simple matter to verify that when a material particle moves in a central force field the quantity $r^2\dot{\varphi}$ is conserved. This is one of Kepler's laws which he discovered while studying the motion of the planets of the Solar system (at that time Kepler already employed the tables of the coordinates of the planets on the celestial sphere as a function of time). This conserved quantity can be given clear geometrical meaning. Kepler introduced a convenient notion of areal velocity v as the time rate of change of the area $s(t)$ swept out by the radius vector $r(t)$, i.e. $v = ds(t)/dt$. In terms of the areal velocity Kepler's law can be formulated as follows: the radius vector sweeps out equal areas in equal times; in other words, the areal velocity is constant, $ds(t)/dt = \text{const}$. We can also prove (the proof is omitted here) that this law is one of the formulations of the principle of conservation of angular momentum. The reader can easily see that this law is much simpler to derive in polar coordinates rather than in Cartesian coordinates (though calculations may, of course, be performed in the latter as well).

Solution of particular problems in mechanics and physics has called for the invention of other curvilinear coordinate systems: cylindrical, spherical, and so on. Close examination of all the aforementioned ways of associating a point in space with a set of real numbers (the coordinates of this point) shows that this association relies upon a general idea admitting a reasonable formalization which comprises all the "curvilinear" coordinates mentioned (inverted commas are used for the word curvilinear, since we have not yet defined the concept strictly, but consider only graphic examples)

1.1.2. CARTESIAN AND CURVILINEAR COORDINATES

Let us consider an arbitrary domain in a Euclidean space $\mathbf{R}^n$. We recall that, just as in mathematical analysis, by a domain we mean an arbitrary set C in a Euclidean space whose every point P is contained in C together with a ball of sufficiently small radius with centre at P. Consider also a second copy of the Euclidean space, which is denoted by $\mathbf{R}_1^n$. To define the coordinates of the point P in the domain C is to associate with this point a set of numbers, called coordinates. Obviously, an arbitrary association (i.e. the one without additional requirements) will not lead to a good result in that such a correspondence may be devoid of sense (it is desirable that mathematical concepts should be of some use, for instance in computations, just as was the case with Cartesian coordinates). Here is an example of senseless association: the same set of numbers, say $(0, 0, 0, \ldots, 0)$, is associated with each point P in C. Thus, we arrive at the first requirement to the association: it is desirable that distinct sets of numbers (coordinates) should correspond to different points of the domain. The example just mentioned does not satisfy this requirement (all the points of the domain C have the same "coordinates", zeros).

Thus, our aim is to associate with each point P of a domain C a set of n real numbers. Apparently, this operation gives rise to a set of n functions $x^1(P), \ldots, x^n(P)$ defined in the domain C; here $x^1, \ldots, x^n$ are coordinates in the Euclidean space $\mathbf{R}_1^n$.

These functions are usually required to be continuous and even smooth (at least for almost all the points of the domain C), that is, a small change in the position of P should lead to a small change in its coordinates, and a smooth deformation of P should generate a smooth variation of its coordinates.

So, let us consider two copies of a Euclidean space: $\mathbf{R}^n$ with Cartesian coordinates $y^1, \ldots, y^n$ and $\mathbf{R}_1^n$ with Cartesian coordinates $x^1, \ldots, x^n$; let C be a domain in $\mathbf{R}^n$.

Remark. The Euclidean space $\mathbf{R}_1^n$ could be considered as an "arithmetic space" by identifying its points with real sequences of length n.

Definition 1. A *continuous coordinate system* in a domain C of Euclidean space $\mathbf{R}^n$ is said to be a system of functions $x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)$ which map the domain C continuously and bijectively onto a certain domain A of $\mathbf{R}_1^n$. In other words, the system of functions $x^1(P), \ldots, x^n(P)$ defines a mapping, sometimes called a homeomorphism of C onto A (the concept of a homeomorphism will be considered in detail below).

Definition 1 is a formal expression of our desire that as a point P moves continuously in C its coordinates should also change continuously. The functions $x^1(P), \ldots, x^n(P)$ are called the *coordinates* of point P relative to the coordinate mapping $f\colon C \to A$.

For instance, the coordinate mapping $f: C \to A$ may be chosen in the form of an identity mapping defined by the linear functions, $x^1 = y^1, \ldots, x^n = y^n$.

Sometimes we shall write a point P with coordinates $x^1(P), \ldots, x^n(P)$ in the form $P(x^1, \ldots, x^n)$ assuming that the coordinate mapping $f: C \to A$ has already been defined and fixed.

Among all continuous coordinate mappings of special interest are those that define a smooth mapping of a domain C onto A, i.e. when all functions $x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)$ are continuous functions of their arguments $y^1, \ldots, y^n$. But the smoothness of the coordinate mapping f without the assumption of the smoothness of the inverse mapping f^{-1} does not lead to a meaningful coordinate system. Therefore, we now turn to defining coordinate systems in which f and f^{-1} are both smooth. To this end, we shall need a new concept, the Jacobi matrix of a smooth mapping.

Let $f: C \to A$ be a smooth mapping defined by a set of functions $x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)$.

Definition 2. The *Jacobi matrix* of a mapping f is a functional matrix

$$df = \left(\frac{\partial x}{\partial y}\right) = \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \cdots & \frac{\partial x^1}{\partial y^n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \frac{\partial x^n}{\partial y^1} & \cdots & \frac{\partial x^n}{\partial y^n} \end{pmatrix},$$

composed of partial derivatives of the coordinates $x^1(P), \ldots, x^n(P)$. The determinant of this matrix is denoted by $J(f)$ and called the *Jacobian* of the mapping f.

Remark. We hope that the notation df for the Jacobi matrix will not be confused with the differential of a smooth function f, for this differential (when interpreted appropriately) coincides with the Jacobi matrix in this particular case. This topic is discussed below. Let us note once more that a Jacobi matrix is a variable matrix, i.e. it depends on a point P in a domain C; similarly, the Jacobian $J(f)$ is a smooth function on C.

Definition 3. A *regular coordinate system* in a domain C of Euclidean space $\mathbf{R}^n$ is a system of smooth functions $x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)$ which map bijectively the domain C onto a domain A in $\mathbf{R}_1^n$ and are such that the Jacobian $J(f)(P)$ is not zero at all points of C.

Let us note that the condition that the Jacobian does not vanish at all points of C means that the inverse mapping f^{-1} is not only continuous, but also smooth. This follows from the implicit-function theorem. Thus, a regular coordinate system is defined by two smooth mutually inverse mappings establishing a homeomor-

phism between the domains C and A. Definition 3 makes formal our desire that when a point P changes smoothly in C its coordinates should also change smoothly; moreover, smooth variation of a "coordinate point" B in A should also result in smooth variation of the point P induced by the mapping f^{-1}. The definitions presented above clearly show that the very concept of a "smooth and regular coordinate system" automatically implies that at least two copies of a standard Euclidean space should be considered. Certain domains of these copies are identified by a continuous and bijective mapping with an additional requirement of smoothness (in both directions).

These definitions can be interpreted from another point of view. We could assume that a Cartesian coordinate system is initially introduced in a domain C of Euclidean space $\mathbf{R}^n$ (via an identity mapping of C onto A under natural identification of both copies, $\mathbf{R}^n$ and $\mathbf{R}^n_1$). Then, the introduction in C of another coordinate system defined by a regular mapping f (i.e. a smooth, one-to-one mapping with a non-zero Jacobian) may be considered as a coordinate transformation: we simply pass from the initial Cartesian coordinate system to a new one in the same domain C.

Definition 4. A regular coordinate system in a domain C is sometimes called a *curvilinear coordinate system* in C.

Consider two arbitrary curvilinear coordinate systems in a domain C: $x^1(P), \ldots, x^n(P)$ and $z^1(P), \ldots, z^n(P)$. This means that two regular mappings $f\colon C \to A \subset \mathbf{R}^n_1(x^1, \ldots, x^n)$ and $g\colon C \to B \subset \mathbf{R}^n_2(z^1, \ldots, z^n)$ are defined which map smoothly and bijectively the domains C, A and C, B, respectively. In other words, each point P in C is associated with two sets of curvilinear coordinates $\{x^i(P)\}$ and $\{z^i(P)\}$, $1 \leqslant i \leqslant n$. Since this correspondence is bijective, we may consider a correspondence which relates the coordinates $\{x^i(P)\}$ of point P to the coordinates $\{z^i(P)\}$, this operation defining the mapping $\psi_{x,z}\colon A \to B$, i.e. $\psi_{x,z}\colon x^i(P) \to z^i(P)$, $1 \leqslant i \leqslant n$. The mapping $\psi_{x,z}$ is called *coordinate transformation* in the domain C. Under this transformation the initial curvilinear coordinates $\{x^i(P)\}$ of point P change to new curvilinear coordinates $\{z^i(P)\}$.

Lemma 1. *The transformation $\psi_{x,z}$ is a bijective and smooth mapping of the domain A onto B with a non-zero Jacobian.*

Proof. That $\psi_{x,z}$ is one-to-one directly follows from Definition 3. The smoothness of $\psi_{x,z}$ follows from the fact that the composition of two smooth mappings is also a smooth mapping. It remains to verify that the Jacobian $J(\psi_{x,z})$ of $\psi_{x,z}$ is non-zero at each point of the domain B.

Indeed, the mapping $\psi_{x,z}$ splits into the composition of two mappings: $\psi_{x,z} = g \circ f^{-1}\colon A \to B$ (Fig. 1.5). The Jacobi matrix of $\psi_{x,z}$ splits into the product of Jacobi matrices of the mappings f^{-1} and g. Indeed, $d\psi_{x,z} = \left(\frac{dz}{dx}\right)$. Consider the derivative $\frac{\partial z^i}{\partial x^j}$; since

$z^i = z^i(y^1, \ldots, y^n) = z^i(y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n))$, where the functions $\{y^\alpha(x^1, \ldots, x^n), 1 \leqslant \alpha \leqslant n\}$ define the smooth mapping $f^{-1}: A \to C$, we obtain from the formula for differentiation of a composite function $\frac{\partial z^i}{\partial x^j} = \sum_{k=1}^{n} \frac{\partial z^i}{\partial y^k} \frac{\partial y^k}{\partial x^j}$, which means that the Jacobi matrix $d\psi_{x,z}$ splits into the product of two matrices dg and df^{-1}. We have used here the formula which expresses the elements of the product of two matrices in terms of the elements of each matrix.

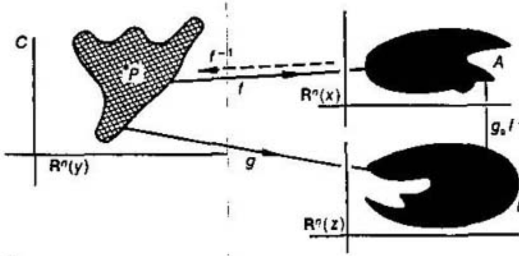

Figure 1.5

It remains to establish a relation between the Jacobi matrices df and $d(f^{-1})$. Since the composition $f^{-1} \circ f$ is an identity mapping of the domain C into itself (see the definition of a regular coordinate system), we find from the formula just proved that $d(f^{-1} \circ f) = df^{-1} \circ df = E$, where E is an identity matrix of order n, i.e. finally $d(f^{-1}) = (df)^{-1}$. Thus, we have proved that the identity $d\psi_{x,z} = (dg) \cdot (df)^{-1}$ holds true for the matrix $d\psi_{x,z}$, i.e. $J(\psi_{x,z}) = J(g)/J(f)$, and since the Jacobians $J(g)$ and $J(f)$ are both non-zero, $J(\psi_{x,z})$ is also non-zero. The lemma is proved.

If the mapping $f: C \to A$ defines curvilinear coordinates in C, the mapping $f^{-1}: A \to C$ defines curvilinear coordinates in A (through Cartesian coordinates in C). We shall often use this simple remark, when passing from the mapping f to f^{-1}.

Let a set of smooth functions $\{x^i(P)\}$, $1 \leqslant i \leqslant n$, be given on a domain C. A question arises: does this set define a regular coordinate system in C?

Lemma 2. *Let the set of smooth functions $\{x^i(P)\}$, $1 \leqslant i \leqslant n$, be such that the Jacobian of this system of functions $J(f = \{x^i(P), 1 \leqslant i \leqslant n\})$ is non-zero in the domain C. Then, for each point P in C there exists an open neighbourhood such that $\{x^i(P)\}$ defines in this neighbourhood a regular coordinate system (such a coordinate system may be called a local coordinate system).*

Proof. The lemma does not presuppose that the set of functions $\{x^i(P)\}$ defines (at least locally) a one-to-one mapping of the domain C onto a domain A of Euclidean space $\mathbf{R}_1^n$. Using the implicit-function theorem (and the existence theorem for inverse mapping), we see that a non-zero Jacobian implies the existence (at least in an open neighbourhood) of an inverse mapping which is also smooth. Thus, the proof of the lemma follows from the definition of a regular coordinate system.

Let us note that the set of functions satisfying Lemma 2 may not define a regular coordinate system in the whole domain C, i.e. the smooth mapping f^{-1} of the domain A onto C may not exist. Here is

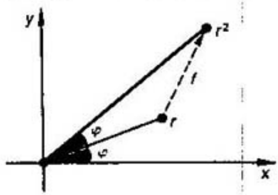

Figure 1.6

a simple example. Let a two-dimensional Cartesian plane with punctured origin O be chosen as the domain C and let the mapping f (defined by two functions $x^1(P)$ and $x^2(P)$) represent the smooth mapping $f(y^1, y^2) = (x^1(y), x^2(y))$, where $x^1(y^1, y^2) = (y^1)^2 - (y^2)^2$ and $x^2(y^1, y^2) = 2y^1y^2$, i.e. if we put $z = y^1 + iy^2$, $w = x^1 + ix^2$ (i is the imaginary unit), then $w = z^2$. This mapping transforms a complex number z into this number squared (for convenience, one may assume that the two copies of the Euclidean plane, $\mathbf{R}^2(y)$ and $\mathbf{R}^2(x)$, are identified with each other). The same mapping can easily be written in the polar coordinates (r, φ) to give $f(r, \varphi) = (r^2, 2\varphi)$. Geometric interpretation of the mapping f is shown in Fig. 1.6. Let us find the Jacobian $J(f)$ (we shall calculate it, for example, in the initial Cartesian coordinate system y^1, y^2 on $\mathbf{R}^2(y)$). The Jacobi matrix df is of the form

$$df = \begin{pmatrix} 2y^1 & 2y^2 \\ -2y^2 & 2y^1 \end{pmatrix}^T, \quad \text{i.e. } J(f) = 4((y^1)^2 + (y^2)^2) > 0.$$

We see that the Jacobian is positive at all points of C (since the origin is punctured). Hence, according to Lemma 2, our mapping establishes a local (regular) coordinate system in an open neighbourhood of each point in C. At the same time, the mapping f does not have the inverse mapping f^{-1} because f is not bijective. Indeed, every point $w = x^1 + ix^2 \in \mathbf{R}^2(x)$, other than the origin, always has exact-

ly two inverse images under the mapping f: namely, the points (r, φ) and $(r, \varphi + \pi)$ which are, of course, distinct points of the domain C. Thus, if a set of functions is chosen as a regular coordinate system in a Euclidean domain C, we should not only verify that the Jacobian of this system is non-zero (at every point of C), but also that the mapping defined by this set is bijective. Note also that in the above example the Jacobian of the system of functions vanishes as the point P tends to zero. A question which often arises in geometry is: will a smooth mapping of a Euclidean space into itself

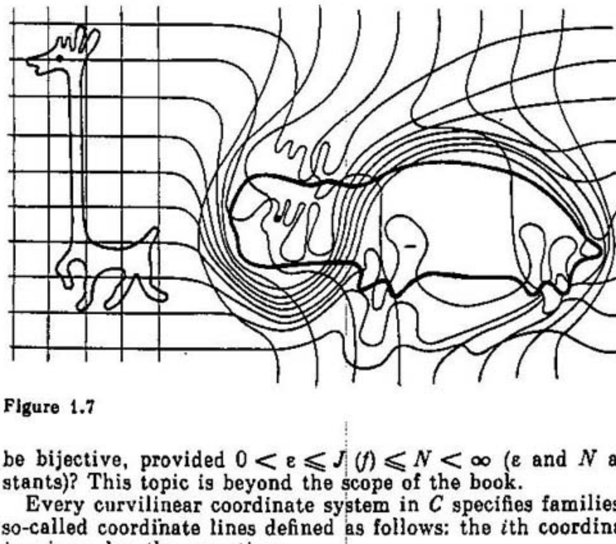

Figure 1.7

be bijective, provided $0 < \varepsilon \leqslant J(f) \leqslant N < \infty$ (ε and N are constants)? This topic is beyond the scope of the book.

Every curvilinear coordinate system in C specifies families of the so-called coordinate lines defined as follows: the ith coordinate line is given by the equations

$$x^1(P) = c_1,\ x^2(P) = c_2,\ \ldots,\ x^{i-1}(P) = c_{i-1},\ x^i(P) = t,$$
$$x^{i+1}(P) = c_{i+1},\ \ldots,\ x^n(P) = c_n,$$

where all c_i are constants and t is a continuous parameter. As t varies, the point P runs a smooth trajectory in the domain C. Thus, n smooth trajectories emerge from each point P in C, and these trajectories are called coordinate lines of a given coordinate system (at point P). For another point P we obtain another family of coordinate lines which are smoothly deformed as point P varies. For example, if we consider a Cartesian coordinate system, the coordinate lines are straight lines through P parallel to the coordinate axes.

In graphic representation of a curvilinear coordinate system it is often useful to draw coordinate lines; in particular, transformation to a curvilinear coordinate system is especially clear if the coordinate network is depicted (see Fig. 1.7).

1.1.3. SIMPLE EXAMPLES OF CURVILINEAR COORDINATE SYSTEMS

We start this section by noting that a polar coordinate system (r, φ) on a Euclidean plane is not a regular coordinate system defined on the whole plane $\mathbf{R}^2$. Indeed, let us consider the functions of coordinate transformation from a polar system to a Carte-

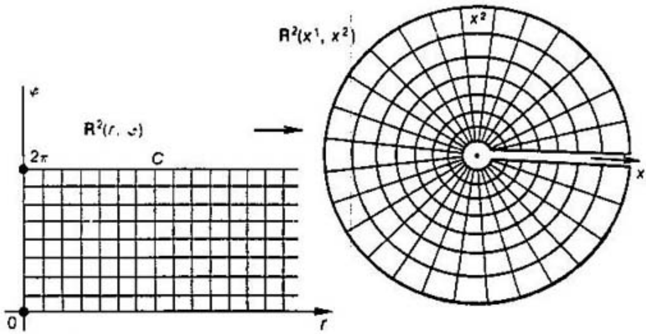

Figure 1.8

sian one: $x^1 = r\cos\varphi$ and $x^2 = r\sin\varphi$. Direct calculation of the transformation Jacobian $J(\psi)$ yields

$$d\psi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix}^T, \quad J(\psi) = r.$$

Thus, since the Jacobian is zero at the origin and the Cartesian coordinate system x^1, x^2 is, obviously, regular on the plane $\mathbf{R}^2(x^1, x^2)$, the polar coordinate system is not regular (in the sense of Definition 3). Furthermore, this is not the only disadvantage of our polar coordinate system. This system is not a bijective mapping of the whole Euclidean plane onto itself, since the points (r, φ) and $(r, \varphi + 2\pi)$ are transformed into the same point.

A domain C in which a polar coordinate system is regular deserves special attention. Let us analyse this example in greater detail.

Consider a Euclidean plane $\mathbf{R}^2(r, \varphi)$, where $y^1 = r$, $y^2 = \varphi$, and take an infinite strip defined by $0 < \varphi < 2\pi$, $0 < r < +\infty$

as a domain C. Then, the domain A in the plane $\mathbf{R}^1(x^1, x^2)$ should be chosen as the entire plane except the ray $x^1 \geqslant 0$, $x^2 = 0$. The mapping $f\colon C \to A$ is given by the relations $x^1 = r\cos\varphi$ and $x^2 = r\sin\varphi$. Figure 1.8 shows how the coordinate lines are trans-

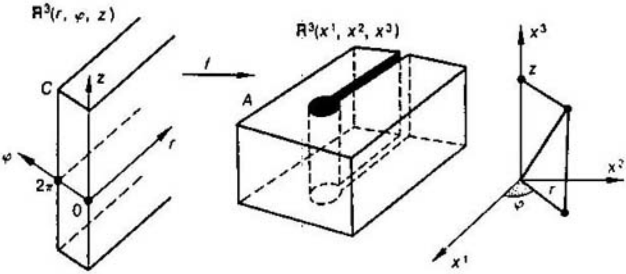

Figure 1.9

formed under this mapping. A rectangular network of Cartesian coordinates goes over to a polar network. Obviously, the mapping f is bijective and regular.

We now turn to a three-dimensional Euclidean space and consider so-called cylindrical coordinates. Coordinate transformation is described by the relations: $x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$, and $x^3 = z$. Consider $\mathbf{R}^3(y^1, y^2, y^3)$, where $y^1 = r$, $y^2 = \varphi$, $y^3 = z$, and take for C the domain $(0 < r,\ 0 < \varphi < 2\pi,\ -\infty < z < +\infty)$. These relations define a smooth mapping $f\colon C \to A \subset \mathbf{R}^3(x^1, x^2, x^3)$, where the domain A is obtained from $\mathbf{R}^3(x^1, x^2, x^3)$ by the elimination of the half-plane (Fig. 1.9). The Jacobi matrix is of the form

$$\begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -r\sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}^T$$

and the transformation Jacobian is equal to r. Thus, in the domain A a cylindrical coordinate system is regular (the Jacobian is zero only at points of the z axis); the half-plane $(\varphi = 0,\ r \geqslant 0)$ is eliminated to make the mapping bijective.

Let us now turn to an n-dimensional Euclidean space and introduce in it a spherical coordinate system. Although the transformation formulas, the Jacobi matrix, and the Jacobian are written for the n-dimensional case, the structure of the domains C and A is analysed

(for clarity) only for the three-dimensional case. The transformation formulas are as follows:

$$f_n\colon C(r, \theta_1, \theta_2, \dots, \theta_{n-1}) \to A(x^1, x^2, \dots, x^n).$$

$$\begin{cases} x^1 = r\cos\theta_1 \\ x^2 = r\sin\theta_1\cos\theta_2 \\ x^3 = r\sin\theta_1\sin\theta_2\cos\theta_3 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x^{n-1} = r\sin\theta_1\sin\theta_2\dots\sin\theta_{n-2}\cos\theta_{n-1} \\ x^n = r\sin\theta_1\sin\theta_2\dots\sin\theta_{n-2}\sin\theta_{n-1} \end{cases}$$

Remark. It is clear from these formulas that they all have the same structure, but the parameters θ_i, starting with the number $i = n$, are equal to zero. The Jacobi matrix is of the form

$$df_n = \begin{pmatrix} \cos\theta_1 & -r\sin\theta_1 & 0 & \dots \\ \sin\theta_1\cos\theta_2 & r\cos\theta_1\cos\theta_2 & -r\sin\theta_1\sin\theta_2 & \dots \\ \vdots & \vdots & \vdots & \\ \sin\theta_1\dots\sin\theta_{n-2}\cos\theta_{n-1} & r\cos\theta_1\dots\sin\theta_{n-2}\cos\theta_{n-1} & 0 & \dots \\ \sin\theta_1\dots\sin\theta_{n-2}\sin\theta_{n-1} & r\cos\theta_1\dots\sin\theta_{n-2}\sin\theta_{n-1} & & \end{pmatrix}.$$

$$\begin{aligned} J_n &= r\sin\theta_1\dots\sin\theta_{n-2}\cos\theta_{n-1}\cos\theta_{n-1}J_{n-1} \\ &\quad + r\sin\theta_1\dots\sin\theta_{n-2}\sin\theta_{n-1}\sin\theta_{n-1}J_{n-1} \\ &= r\sin\theta_1\dots\sin\theta_{n-2}J_{n-1}, \end{aligned}$$

$J_2 = r$, $J_3 = +r^2\sin\theta_1$, $J_4 = r^3\sin^2\theta_1\sin\theta_2$,
$J_5 = r^4\sin^3\theta_1\sin^2\theta_2\sin\theta_3$ and so on.

Spherical coordinates in a three-dimensional Euclidean space are usually denoted by (r, θ, φ); in this notation the transformation formulas become $x^1 = r\sin\theta\cos\varphi$, $x^2 = r\sin\theta\sin\varphi$, $x^3 =$

$r\cos\theta$, $0 < \theta < \pi$, $0 < \varphi < 2\pi$, $r \geqslant 0$. In these coordinates $J = r^2 \sin\theta$. Consider the structure of domains C and A (see Fig. 1.10). The Jacobian of this transformation vanishes only at points of the

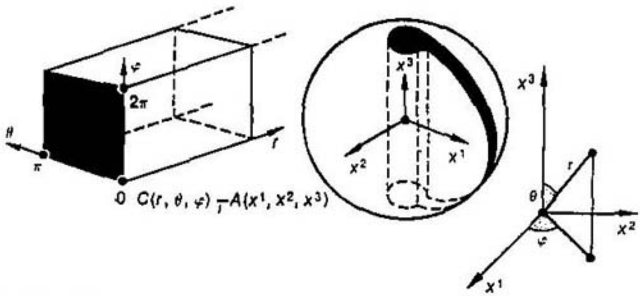

Figure 1.10

x^3 axis, the remaining points belonging to the half-plane ($x^2 = 0$, $x^1 \geqslant 0$) are eliminated to make it possible to introduce a bijective coordinate system. For a fixed r, the coordinate lines of the para-

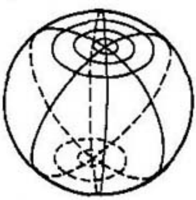

Figure 1.11

meters θ and φ are shown in Fig. 1.11. These angular parameters are sometimes called latitude and longitude (they provide a coordinate network on a globe). The Jacobi matrix in the three-dimensional case is

$$d\psi = \begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}.$$

Problems

1. Prove that the system of functions $u = x + \sin y$, $v = y - \frac{1}{2}\sin x$ on a plane is a regular coordinate system.

2. Demonstrate that a global coordinate system cannot be defined on a circle S^1.

3. Write the Laplace equation $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ in a polar coordinate system.

1.2. THE LENGTH OF CURVE IN A CURVILINEAR COORDINATE SYSTEM

1.2.1. THE LENGTH OF CURVE IN A EUCLIDEAN COORDINATE SYSTEM

Consider a Euclidean space $\mathbf{R}^n$ and define in it a Euclidean scalar product $\langle \xi, \eta \rangle = \Sigma \xi^i \eta^i$, $\xi, \eta \in \mathbf{R}^n$. Then, with each vector $\xi \in \mathbf{R}^n$ we can associate a real number called the absolute value or length and defined as $|\xi| = \sqrt{\langle \xi, \xi \rangle}$. This formula gives the length of a vector emerging from point 0 to a point $\xi \in \mathbf{R}^n$. To find the distance between any two points $\xi, \eta \in \mathbf{R}^n$, we have to calculate the length of the vector $\xi - \eta$. It is known from analytic geometry that the angle φ between two vectors $\xi, \eta \in \mathbf{R}^n$ can also be expressed in terms of the scalar product as $\cos \varphi = \frac{\langle \xi, \eta \rangle}{|\xi| \cdot |\eta|}$. We thus see that such important metric concepts as the length of a vector and the angle between two vectors are closely related to the Euclidean scalar product. While defining other important geometric concepts, we shall often proceed from the scalar product of vectors.

We know how to calculate the length of a straight segment; it would, however, be desirable to know how to calculate the length of a smooth curve. Let us define the length of a smooth curve $\gamma(t)$, assuming that it is given in parametric form, i.e. by a set of n smooth functions $x^1(t), \ldots, x^n(t)$ in a Euclidean space, where the parameter (time) runs either the entire real axis or the segment $[a, b]$ (the latter case is of major importance to us). We shall also assume that $x^1, \ldots, x^n$ are Cartesian coordinates in $\mathbf{R}^n$.

Definition 1. The *length of a curve* $\gamma(t)$ from point $\gamma(a)$ to point $\gamma(b)$ (or from the parameter value $t = a$ to $t = b$) is the number $l(\gamma)_a^b = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle}\, dt$, where $\dot{\gamma}(t)$ is the vector with the coordinates $\left(\frac{dx^1(t)}{dt}, \ldots, \frac{dx^n(t)}{dt}\right)$ called sometimes the *velocity vector* of the curve $\gamma(t)$ at point t or the *tangent vector* to the curve $\gamma(t)$.

Thus, the length of a curve is the integral of the length of the velocity vector (or of the velocity vector magnitude). The explicit expression for the length of a curve is

$$l(\gamma)_a^b = \int_a^b \sqrt{\sum_{i=1}^n \left(\frac{dx^i(t)}{dt}\right)^2}\, dt.$$

Lemma 1. *Given a smooth curve $\gamma(t)$ and two fixed points $\gamma(a)$ and $\gamma(b)$ corresponding to the parameter values $t = a$ and $t = b$. Let $t = t(\tau)$ be an arbitrary smooth transformation of the parameter t into a new one τ such that $\frac{dt}{d\tau} > 0$. Then, the length of the curve $l(\gamma(t))_a^b$ remains unchanged, i.e. the equality $l(\gamma(t))_a^b = l(\gamma(\tau))_\alpha^\beta$ holds, where $a = t(\alpha)$, $b = t(\beta)$.*

Proof. Straightforward calculation yields

$$l(\gamma(t))_a^b = \int_a^b \sqrt{\langle\dot\gamma_t, \dot\gamma_t\rangle}\, dt = \int_\alpha^\beta \sqrt{\langle\dot\gamma_\tau, \dot\gamma_\tau\rangle\left(\frac{d\tau}{dt}\right)^2}\, \frac{dt}{d\tau}\, d\tau$$
$$= \int_\alpha^\beta \sqrt{\langle\dot\gamma_\tau, \dot\gamma_\tau\rangle}\, \frac{d\tau}{dt}\frac{dt}{d\tau}\, d\tau = \int_\alpha^\beta \sqrt{\langle\dot\gamma_\tau, \dot\gamma_\tau\rangle}\, d\tau,$$

which is what was required. Here we have introduced the notation $\dot\gamma_t = \left\{\frac{dx^1(t)}{dt}, \ldots, \frac{dx^n(t)}{dt}\right\}$.

Let us assume that we have two smooth curves $\gamma_1(t)$ and $\gamma_2(\tau)$ intersecting at a point P of a Euclidean space, that is, there exist parameter values $t = a$ and $\tau = b$ such that $P = \gamma_1(a) = \gamma_2(b)$. Let us define the angle between the two curves at the intersection point.

Definition 2. The *angle* between two smooth trajectories $\gamma_1(t)$ and $\gamma_2(\tau)$ intersecting at point $P = \gamma_1(a) = \gamma_2(b)$ is defined as $\cos\varphi = \frac{\langle\dot\gamma_1(a), \dot\gamma_2(b)\rangle}{|\dot\gamma_1(a)|\cdot|\dot\gamma_2(b)|}$, provided the velocity vectors $\dot\gamma_1(a)$ and $\dot\gamma_2(b)$ are both non-zero at P.

Remark. Strictly speaking, this formula defines not one, but two angles equal together to π. If we assume however that the curves are numbered, we arrive at the concept of an oriented angle which is uniquely defined by the above formula. Consider the condition that the two velocity vectors be nonzero at the intersection point. At the points where the velocity vector of a smooth curve vanishes the curve may have a cusp and abruptly change its direction. Therefore, there is an uncertainty in defining the angle at the point of intersection with such a curve, in regard to the choice of one of the two

smooth segments of the curve separated by the cusp (Fig. 1.12). It should also be noted that the existence of a cusp on a smooth curve (for those parameter values for which the velocity vector vanishes) by no means contradicts the condition that the curve is smooth (see

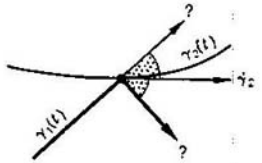

Figure 1.12

the definition of a smooth curve). Figure 1.13 gives an example of a smooth curve with a cusp; here the "cusp angle" at the singular point is equal to $\pi/2$. It is a simple matter to invent a smooth curve with the cusp angle at the singular point equal to π (see Fig. 1.14). (Exercise. Write a parametric equation of the curve of Fig. 1.14. Question: can this curve be defined by analytic functions $x(t)$ and $y(t)$?)

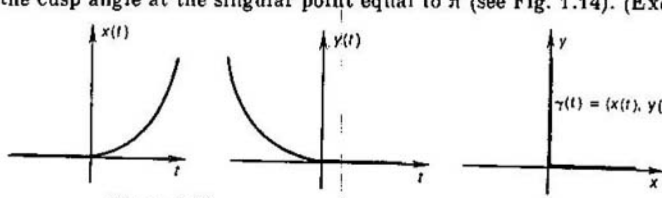

Figure 1.13

Let us now discuss to what extent the above definition of the length of a curve is in conformity with the customary ideas about

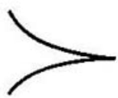

Figure 1.14

length based on the concept of the Euclidean length of a segment and on the concept of a thin inextensible string with the aid of which we can measure the length of a more complicated curve. We know that the circumference of a circle is sometimes defined as the limit of the perimeter of a polygon inscribed in (or circumscribed about)

the circle when the number of the polygon sides tends to infinity. We start with a segment, of course.

(1) Let the curve $\gamma(t)$ be defined by the linear functions $x^i(t) = \alpha^i t$, where $\alpha^i = \text{const}$, $1 \leqslant i \leqslant n$, $a \leqslant t \leqslant b$. Then, calculation of the length of this smooth curve from $t = a$ to $t = b$ yields

$$l(\gamma)_a^b = \int_a^b \sqrt{\sum_{i=1}^n (\alpha^i)^2}\, dt = (b-a)\sqrt{\sum_{i=1}^n (\alpha^i)^2}.$$

Since the segment originates at the point $\{\alpha^i a\}$ and terminates at the point $\{\alpha^i b\}$, the ordinary (Euclidean) length of the segment is $(b-a)\sqrt{\sum_{i=1}^n (\alpha^i)^2}$, which coincides with the value of the integral $l(\gamma)_a^b$.

(2) Let the curve $\gamma(t)$ be defined on the plane $\mathbf{R}^2(x, y)$ by the parametric equations $x = R\cos t$, $y = R\sin t$. Straightforward evaluation of the integral gives $l(\gamma)_a^b = 2\pi R$ (where $a = 0$, $b = 2\pi$), which, apparently, coincides with the familiar expression for the length of a circle of radius R.

1.2.2. THE LENGTH OF CURVE IN A CURVILINEAR COORDINATE SYSTEM

Let us now consider an arbitrary curvilinear coordinate system in a Euclidean domain C and let $\gamma(t)$ be an arbitrary smooth curve in this domain. What is the expression for the length of $\gamma(t)$ in this curvilinear coordinate system? To answer this question, we analyse the behavior of the components of the velocity vector of $\gamma(t)$ under coordinate transformation. Denoting the curvilinear coordinates by $z^1, \ldots, z^n$, i.e. $x^i = x^i(z)$, we obtain from the law of differentiation of a composite function

$$\frac{dx^i(t)}{dt} = \frac{dx^i(z(t))}{dt} = \sum_{(k)} \frac{\partial x^i}{\partial z^k}\frac{dz^k}{dt}, \quad 1 \leqslant i \leqslant n,$$

that is,

$$\begin{aligned} l(\gamma)_a^b &= \int_a^b \sqrt{\sum_{i=1}^n \left(\frac{dx^i(z(t))}{dt}\right)^2}\, dt \\ &= \int_a^b \sqrt{\sum_{i=1}^n \left(\sum_{(k)} \frac{\partial x^i(z(t))}{\partial z^k}\frac{dz^k}{dt}\right)^2}\, dt \end{aligned}$$

$$= \int_a^b \sqrt{\sum_{i=1}^{n} \sum_{m,\,p} \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p} \frac{dz^m}{dt} \frac{dz^p}{dt}}\, dt$$

$$= \int_a^b \sqrt{\sum_{m,\,p} g_{mp}(z) \frac{dz^m}{dt} \frac{dz^p}{dt}}\, dt,$$

where the functions $g_{mp}(z)$ are of the form $g_{mp}(z) = \sum_{i=1}^{n} \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p}$. Obviously, these functions are symmetric in m and p, i.e. $g_{mp} = g_{pm}$. Hence, the set of functions g_{mp} can be represented in the form of a symmetric matrix denoted in the sequel by $G = (g_{mp})$. In the formula just derived the coefficients of G are expressed as the sums of the products of elements of the Jacobi matrices. Indeed, $d\psi_{z,x} = \left(\frac{\partial x}{\partial z}\right)$, so that G is the product of two matrices $G = A^T \cdot A$, where $A = d\psi_{z,x}$. Clearly, the matrix G depends on the curvilinear coordinate system z and may, in general, vary under coordinate transformation. What is the law of variation of $G(z)$? Let us make another coordinate transformation $\{z^i\} \to \{y^k\}$, that is, consider a regular coordinate transformation $z^i = z^i(y^1, \ldots, y^n)$, $1 \leqslant i \leqslant n$. (The set $\{y^k\}$ is again considered as curvilinear coordinates in C.)

In this case the coefficients $g_{mp}(z)$ are transformed as

$$g_{kl}(y) = \sum_{i=1}^{n} \frac{\partial x^i(y)}{\partial y^k} \frac{\partial x^i(y)}{\partial y^l} = \sum_{i=1}^{n} \sum_{m,\,p} \frac{\partial x^i}{\partial z^m} \frac{\partial z^m}{\partial y^k} \frac{\partial x^i}{\partial z^p} \frac{\partial z^p}{\partial y^l}$$

$$= \sum_{m,\,p} \frac{\partial z^m}{\partial y^k} \left(\sum_{i=1}^{n} \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p} \right) \frac{\partial z^p}{\partial y^l} = \sum_{m,\,p} \frac{\partial z^m}{\partial y^k} g_{mp}(z) \frac{\partial z^p}{\partial y^l},$$

i.e. $G(y) = (d\psi_{y,z})^T G(z)\,(d\psi_{y,z})$.

Remark. To simplify the notation, in expressions of the form $\sum_{(i)} a^i b^i$, where summation extends over repeated indices, we shall omit the symbol Σ and simply write $a^i b^i$. For example, the formula $g_{kl}(y) = \sum_{i=1}^{n} \frac{\partial x^i(y)}{\partial y^k} \frac{\partial x^i(y)}{\partial y^l}$ is written as $g_{kl}(y) = \frac{\partial x^i(y)}{\partial y^k} \cdot \frac{\partial x^i(y)}{\partial y^l}$. Sometimes, we shall denote new coordinates by $z^{i'}$, using primed indices for a new coordinate system, viz., $g_{i'j'}$.

The functions $g_{mp}(z)$ have a clear geometric meaning. Consider an arbitrary point P in a domain C and coordinate lines of the curvilin-

ear coordinate system $z^1, \ldots, z^n$ through P. Each of these lines can be defined by the following parametric equations: $z^1 = c_1, \ldots, z^{i-1} = c_{i-1}, z^i = t, z^{i+1} = c_{i+1}, \ldots, z^n = c_n$, where c_α $(1 \leqslant \alpha\ (\neq i) \leqslant n)$ are constants such that P has the coordinates (in the system z): $\{z^\alpha = c_\alpha, 1 \leqslant \alpha \leqslant n\}$. Denote the mth coordinate line by $\gamma_m(t)$, $1 \leqslant m \leqslant n$. Then, in the coordinate system x the mth coordinate line of the system z is written as $\{x^i(c_1, \ldots, c_{m-1}, z^m = t, c_{m+1}, \ldots, c_n)\}$, $1 \leqslant i \leqslant n$.

The velocity vector of this smooth curve at point P has the coordinates $\mathbf{e}_m = \left\{\frac{\partial x^i}{\partial z^m}\right\}$, $1 \leqslant i \leqslant n$. Since $g_{mp}(z) = \sum\limits_i \frac{\partial x^i}{\partial z^m}\frac{\partial x^i}{\partial z^p}$, we can write $g_{mp}(z) = \langle \mathbf{e}_m, \mathbf{e}_p \rangle$, i.e. the functions g_{mp} are scalar products of the vectors tangent to the corresponding coordinate lines

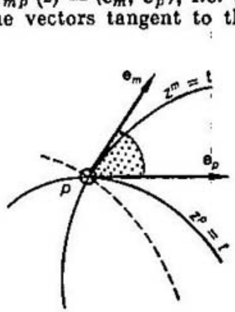

Figure 1.15

(Fig. 1.15) (we recall that the scalar product is Euclidean, i.e. it is a characteristic of the surrounding space).

Thus we see that under coordinate transformation the matrix $G(z)$ is transformed as the matrix of a quadratic form. In particular, if the initial coordinates are Cartesian, G is an identity matrix, so that in any other (curvilinear) coordinate system it can be written as $G(z) = A^T E A$, where $A = d\psi_{z,x}$, provided $\{x^i\}$ are Cartesian coordinates (then $G(x) = E$).

Before proceeding further, we shall consider formulas for the length of a curve in various curvilinear coordinate systems and present explicit expressions for the matrix G in these coordinate systems.

(1) Polar coordinates on a two-dimensional plane $\mathbf{R}^2(r, \varphi)$. In Cartesian coordinates (x^1, x^2) the matrix $G(x)$ is of the form $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, i.e. $g_{ij} = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$. The Jacobi matrix (calculated

above) is $d\psi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix}^T$, whence

$$G(r, \varphi) = d\psi^T (d\psi)$$
$$= \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\sin\varphi \end{pmatrix} \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

Hence, in a polar coordinate system the length of a curve defined as $\gamma(t) = (r(t), \varphi(t))$ is given by

$$l(\gamma)_a^b = \int_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\varphi}{dt}\right)^2}\, dt.$$

(2) Cylindrical coordinates in a three-dimensional Euclidean space $\mathbb{R}^3(r, \varphi, z)$. The matrix $G(x)$ in Cartesian coordinates (x^1, x^2, x^3) is $G(x) = E$. The Jacobi matrix for this case has been calculated above

$$d\psi = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -r\sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}^T,$$

whence

$$G(r, \varphi, z) = (d\psi)^T (d\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, in a cylindrical coordinate system the length of a curve $\gamma(t) = (r(t), \varphi(t), z(t))$ is of the form

$$l(\gamma)_a^b = \int_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\varphi}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt.$$

As an example, we calculate the length of a helix on a right circular cylinder given parametrically as: $r(t) = R = \text{const}$, $\varphi(t) = \omega t$, $z = qt$. We have then

$$l(\gamma)_a^b = \int_a^b \sqrt{R^2\omega^2 + q^2}\, dt = \sqrt{R^2\omega^2 + q^2}\,(b - a).$$

(3) Spherical coordinates in a three-dimensional Euclidean space $\mathbb{R}^3(r, \theta, \varphi)$. The matrix $G(x)$ in Cartesian coordinates (x^1, x^2, x^3) is $G(x) = E$. The Jacobi matrix (calculated above) is of the form

$$\begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix},$$

whence

$$G(r, \theta, \varphi) = (d\psi)^T (d\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}.$$

Hence, in a spherical coordinate system the length of a curve given by $\gamma(t) = (r(t), \theta(t), \varphi(t))$ is expressed as

$$l(\gamma)_a^b = \int_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + r^2 \sin^2\theta \left(\frac{d\varphi}{dt}\right)^2}\, dt.$$

It is sometimes convenient to deal with an elementary arc dl instead of the whole arc. In the examples considered above these elementary arcs squared (in the respective coordinates) are of the form: $(dl)^2 = (dr)^2 + r^2 (d\varphi)^2$ (polar coordinates on a plane), $(dl)^2 = (dr)^2 + r^2 (d\varphi)^2 + (dz)^2$ (cylindrical coordinates in $\mathbf{R}^3$); $(dl)^2 = (dr)^2 + r^2 (d\theta)^2 + r^2 \sin^2\theta\, (d\varphi)^2$ (spherical coordinates in $\mathbf{R}^3$).

1.2.3. THE CONCEPT OF THE RIEMANNIAN METRIC IN A EUCLIDEAN DOMAIN

In the preceding section we have associated with each curvilinear coordinate system z in C a smooth matrix function $G(z)$ which is transformed under coordinate transformation as a quadratic form (at each point). The role of this set of matrix functions is quite clear: once it is given, the length of a curve can be calculated in a curvilinear coordinate system. Among all the properties of this set of matrix functions of major importance is the property of being transformed (at each point) as a quadratic form. Consider now various sets satisfying this property.

Definition 2. A *Riemannian metric* is said to be given in a Euclidean domain C if in any regular coordinate system $z^1, \ldots, z^n$ there is defined a set of smooth functions $g_{mp}(z^1, \ldots, z^n)$ such that: (1) $g_{mp}(z) = g_{pm}(z)$ ($G(z)$ is symmetric), (2) $G(z) = (g_{mp})$ is non-singular and positive definite, (3) under coordinate transformation $z \to y$ the matrix $G(z)$ is transformed as $G(y) = d\psi^T G(z)\,(d\psi)$ (recall that only regular coordinate transformations are considered). Here $d\psi$ stands for the Jacobi matrix of the coordinate transformation $d\psi_{y,z}$.

Definition 3. Let the Riemannian metric $G(z) = g_{ij}$ be given in C and let a smooth curve $\gamma(t) = \{z^i(t)\}$ be given in the coordinate system (z^i). The *length* of the curve from point $\gamma(a)$ to point $\gamma(b)$ is the number

$$l(\gamma)_a^b = \int_a^b \sqrt{g_{ij}(z)\left(\frac{dz^i(t)}{dt}\right)\left(\frac{dz^j(t)}{dt}\right)}\, dt.$$

If two smooth curves $\gamma_1(t)$ and $\gamma_2(t)$ satisfying $\gamma_1(0) = \gamma_2(0) = P$, $\dot{\gamma}_1(0) \neq 0$, $\dot{\gamma}_2(0) \neq 0$ intersect at a point $P \in C$, the angle between these curves (for a given Riemannian metric) is the number φ such that

$$\cos\varphi = \frac{g_{ij}(z)\frac{dz_1^i(t)}{dt}\frac{dz_2^j(t)}{dt}}{\sqrt{g_{ij}(z)\frac{dz_1^i(t)}{dt}\frac{dz_1^j(t)}{dt}}\sqrt{g_{ij}(z)\frac{dz_2^i(t)}{dt}\frac{dz_2^j(t)}{dt}}}.$$

The definition of the Riemannian metric presented above can be formulated in more invariant terms without explicit coordinate writing of the metric. Namely, once the Riemannian metric has been given, we can define (at each point of a domain) a bilinear form $\langle,\rangle_g$ on the set of all vectors tangent to smooth trajectories through this point. Indeed, if $\gamma_1(0) = P = \gamma_2(0)$, where $P \in C$, then we may assume $\langle\xi, \eta\rangle_g = g_{ij}\xi^i\eta^j$ for the vectors $\xi = \dot{\gamma}_1(0)$ and $\eta = \dot{\gamma}_2(0)$, where $\xi = (\xi^1, \ldots, \xi^n)$, $\eta = (\eta^1, \ldots, \eta^n)$. Note that the coordinates of the vectors ξ and η are calculated in the system $(z^1, \ldots, z^n)$.

Lemma 2. *The mapping $\xi, \eta \to \langle\xi, \eta\rangle_g$ defines a non-degenerate, positive definite bilinear form smoothly dependent on a point.*

Proof. That the mapping is symmetric and bilinear follows from Definition 2. We now verify that the mapping defines a bilinear form. We perform a regular coordinate transformation $\{z^i\} \to \{z^{i'}\}$ and have

$$\gamma_1(t) = \{z_1^1(t), \ldots, z_1^n(t)\}, \quad \gamma_2(t) = \{z_2^1(t), \ldots, z_2^n(t)\},$$

$$\xi^{i'} = \frac{dz_1^{i'}}{dt} = \frac{\partial z^{i'}}{\partial z^i}\frac{dz_1^i}{dt} = \frac{\partial z^{i'}}{\partial z^i}\xi^i, \quad \eta^{i'} = \frac{\partial z^{i'}}{\partial z^i}\eta^i, \quad g_{i'j'} = \frac{\partial z^i}{\partial z^{i'}}\frac{\partial z^j}{\partial z^{j'}}g_{ij}$$

whence

$$\begin{aligned}\langle\xi', \eta'\rangle_{g'} = g_{i'j'}\xi^{i'}\eta^{j'} &= \frac{\partial z^i}{\partial z^{i'}}\frac{\partial z^j}{\partial z^{j'}}g_{ij}\frac{\partial z^{i'}}{\partial z^k}\xi^k\frac{\partial z^{j'}}{\partial z^p}\eta^p \\ &= \left(\frac{\partial z^i}{\partial z^{i'}}\frac{\partial z^{i'}}{\partial z^k}\right)\left(\frac{\partial z^j}{\partial z^{j'}}\frac{\partial z^{j'}}{\partial z^p}\right)g_{ij}\xi^k\eta^p = \delta_k^i\delta_p^j g_{ij}\xi^k\eta^n \\ &= g_{ij}\xi^i\eta^j = \langle\xi, \eta\rangle_g,\end{aligned}$$

i.e. $\langle\xi, \eta\rangle_g$ is really a bilinear form. We have used the relation $\frac{\partial z^i}{\partial z^{i'}}\frac{\partial z^{i'}}{\partial z^k} = \delta_k^i$ which follows from the identity $(d\psi)(d\psi)^{-1} = E$. The lemma is proved.

Thus, we arrive at the following definition: the Riemannian metric

is said to be defined in a Euclidean domain C if at each point of the domain there is given a bilinear form (scalar product) on vectors tangent to smooth curves through this point, the form being nondegenerate, positive definite, symmetric, and smoothly dependent on the point.

Lemma 2 shows that this definition is equivalent to Definition 2. The lemma implies, in particular, that the length of a smooth curve does not depend on a curvilinear coordinate system (provided the Riemannian metric is given in a certain coordinate system and is transformed under coordinate substitution in accordance with the law considered above).

Do Riemannian metrics exist? An example has already been given. Indeed, if the matrix $G(x) = (\delta_{ij})$ is defined in a domain C in the Cartesian coordinates $(x^1, \ldots, x^n)$, then in any other curvilinear coordinate system z obtained from x by a regular transformation we can set, by definition, $G(z) = d\psi^T G(x)(d\psi) = (d\psi)^T(d\psi)$, where $d\psi$ is the Jacobi matrix of this transformation. Apparently, since any regular coordinate system z can be obtained from a Cartesian coordinate system by a regular transformation (see the above definition of a regular coordinate system), we have defined the matrix function $G(z)$ in an arbitrary regular coordinate system. This consideration shows that conditions (1)-(3), which the Riemannian metric must satisfy, hold true. Indeed, properties (1) and (2) can be verified directly by the definition of $G(z)$, property (3) follows from the fact that the Jacobi matrix of the composition of two coordinate transformations is equal to the product of the Jacobi matrices of each transformation: $d\psi_{z_1, z_3} = d\psi_{z_1, z_2} \cdot d\psi_{z_2, z_3}$. Thus, we have defined a Riemannian metric in a Euclidean domain. This metric is Euclidean, and the squared differential of the length of a smooth curve in a Cartesian coordinate system is written as $(dl)^2 = \sum_{i=1}^{n} (dx^i)^2$. However, it should not be thought that, given an arbitrary Riemannian metric, an appropriate coordinate transformation may be found in C which will reduce the metric to the form $\sum_{i=1}^{n} (dx^i)^2$.

Definition 4. A Riemannian metric G defined in a domain C is called *Euclidean* if in C there exists a coordinate system y (generally curvilinear) such that $G(y)$ is an identity matrix.

Given a Euclidean metric (relative to a coordinate system), we can describe all other coordinate systems in which this metric is also Euclidean (there are many such systems). This description will be given in the sequel, for at this moment we do not have at our disposal an appropriate apparatus. Here we only note that all such systems can be obtained from one system by rotations, translations, and reflections in a Euclidean space.

The existence of a "non-Euclidean" metric, i.e. a metric which cannot be represented as $\sum_{i=1}^{n} (dx^i)^2$ in any coordinate system, cannot be demonstrated for the time being. At this moment we are unable to find such a metric about which we could say with certainty that it is non-Euclidean. It is intuitively clear that to find such a metric, one has to find its invariants preserved under a regular coordinate transformation. Once these invariants are found for two metrics and shown to be distinct, we can prove the non-equivalence of the two metrics. Such invariants do exist and we shall define them soon. Only after that shall we be able to prove rigorously the existence of a non-Euclidean metric defined in a Euclidean domain.

As was noted earlier, the Euclidean metric written in an arbitrary coordinate system z "loses" its simple "Euclidean form" and is given by the matrix $G(z)$ in which it is rather difficult to "recognize" this metric, especially if we do not know the invariants discriminating different metrics (that is, the metrics which are not transformed into one another under a suitable coordinate transformation). We now show how the Euclidean metric is written in simple curvilinear coordinate systems considered above. It is often more convenient to deal with the squared elementary arc of a smooth curve $(dl)^2 = g_{ij}(z)\, dz^i dz^j$ rather than with the matrix $G(z)$ of the Riemannian metric.

(1) Polar coordinates on a plane: $dl^2 = dr^2 + r^2\, d\varphi^2$.

(2) Cylindrical coordinates in three-dimensional space: $dl^2 = dr^2 + r^2\, d\varphi^2 + dz^2$.

(3) Spherical coordinates in three-dimensional space: $dl^2 = dr^2 + r^2\,(d\theta^2 + \sin^2\theta\, d\varphi^2)$.

1.2.4. INDEFINITE METRICS

Until now we have considered only positive definite metrics called Riemannian. In various applications, however, we often come across the so-called indefinite metrics.

Definition 5. An *indefinite metric* is said to be given in a Euclidean domain C if in any regular coordinate system $z^1, \ldots, z^n$ there is defined a set of smooth functions $\{g_{mp}(z^1, \ldots, z^n)\}$ satisfying all the conditions imposed on the Riemannian metric (see Definition 2) except the condition of positive definiteness, i.e. the corresponding quadratic form is indefinite.

As a simple example of an indefinite metric, we shall consider the so-called pseudo-Euclidean metric of index s in a pseudo-Euclidean space $\mathbf{R}^n_s$. To construct such a metric, it is sufficient to provide an ordinary Euclidean space $\mathbf{R}^n$ with Cartesian coordinates $x^1, \ldots, x^n$ and define at each point $P \in \mathbf{R}^n$ the following bilinear form (with

constant, i.e. point-independent coefficients): $\langle \xi, \eta \rangle_s = -\sum_{i=1}^{n} \xi^i \eta^i + \sum_{j=s+1}^{n} \xi^j \eta^j$. Then, for any smooth curve $\gamma(t) = \{x^i(t)\}$, $1 \leqslant i \leqslant n$, the length of an arc is expressed as

$$l(\gamma)_a^b = \int_a^b \sqrt{-\sum_{i=1}^{s} \left(\frac{dx^i}{dt}\right)^2 + \sum_{j=s+1}^{n} \left(\frac{dx^j}{dt}\right)^2}\, dt.$$

For $n = 4$, pseudo-Euclidean space of index 1 is sometimes called *Minkowski's space* (in the special theory of relativity); we shall also

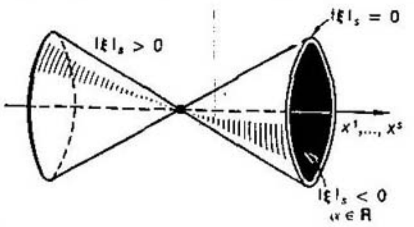

Figure 1.16

consider spaces $\mathbf{R}_1^3$ and $\mathbf{R}_1^2$. Note that pseudo-Euclidean space of index 0 coincides with an ordinary Euclidean space.

Remark. The study of spaces $\mathbf{R}_{n-s}^n$ is equivalent, in a sense, to the study of $\mathbf{R}_s^n$, since all lengths in $\mathbf{R}_{n-s}^n$ can be multiplied by i (the imaginary unit); the form $\langle , \rangle_{n-s}$ changes then to $\langle , \rangle_s$. Therefore, we shall assume, for simplicity, that the inequality $s \leqslant [n/2]$ holds true (square brackets denote the integral part).

As in an ordinary Euclidean space, the length of a vector ξ in $\mathbf{R}_s^n$ is given by $|\xi|_s = \sqrt{\langle \xi, \xi \rangle_s}$, but in $\mathbf{R}_s^n$, unlike in $\mathbf{R}^n$, the length of a vector may be zero or imaginary. Indeed, since the form $\langle , \rangle_s$ is not positive definite, the set of all vectors $\xi \in \mathbf{R}^n$ emerging, say, from the origin can be decomposed into three disjoint subsets: $\langle \xi, \xi \rangle_s < 0$ (*time-like* vectors), $\langle \xi, \xi \rangle_s = 0$ (*light* or *isotropic* vectors), $\langle \xi, \xi \rangle_s > 0$ (*space-like* vectors). Owing to this circumstance, there may exist vectors with zero, real, and purely imaginary length. Indeed, the length of a time-like vector is purely imaginary, that of a light vector is zero, and a space-like vector has a real length. The positions of these three types of vectors in space are also different. Consider the vectors emerging from the origin. The definition implies

that isotropic vectors form the cone: $-\sum_{i=1}^{s}(x^i)^2+\sum_{j=s+1}^{n}(x^j)^2=0$ with the vertex at the origin, time-like vectors are located "inside" the cone, i.e. in its hollow bounded by the coordinate plane $x^1, \ldots, x^s$, and space-like vectors are located "outside" the light cone (Fig. 1.16). This fact in itself shows that the indefinite metric defines a much richer geometry (in the metric sense) than the Euclidean metric.

Remark. In Minknowski's space $\mathbf{R}_1^4$ (in the special theory of relativity) the isotropic cone consists entirely of the so-called "light vectors" ξ (i.e. $\langle\xi, \xi\rangle_1 = 0$) and is termed the light cone, since a light ray emerging from the origin propagates along one of the generators of this cone (provided the parameter ct is chosen as the coordinate x^1, the constant c is the velocity of light).

As in a Euclidean space, we can now define the *length* of any smooth curve in the pseudo-Euclidean space $\mathbf{R}_s^n$ of index s by setting $l_s(\gamma)_a^b = \int_a^b \sqrt{\langle\dot\gamma, \dot\gamma\rangle_s}\, dt$, the only difference from the Euclidean case being that certain curves may have zero, purely imaginary or complex length. Indeed, since the integral along a curve is additive, it can be decomposed into a sum of several terms (for simplicity, their number is assumed to be finite), each term being such that at each point of a given segment of the curve the scalar product $\langle\dot\gamma, \dot\gamma\rangle_s$ does not change the sign (segments of zero length may also occur). Thus, the length of a curve is, in general, a complex number.

The space $\mathbf{R}_1^4$ introduced for conveniently expressing certain quantities in the special theory of relativity has played an important role. Let us refer $\mathbf{R}_1^4$ to the coordinates $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$, i.e. $dl^2 = -c^2\,dt^2 + dx^2 + dy^2 + dz^2$. Here t is time and c is the velocity of light. Consider in $\mathbf{R}_1^4$ an orthonormal 4-frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ with respect to the Euclidean metric in $\mathbf{R}^4$. Here the space $\mathbf{R}_1^4$ is superposed, for the sake of clarity, with the Euclidean four-dimensional space. Consider also the so-called "world line" of a particle $\gamma(\tau)$, this line being a smooth trajectory in $\mathbf{R}_1^4$. If x, y, z are spatial coordinates, the motion of the particle along the trajectory $\gamma(\tau)$ can be interpreted as the evolution in space and time of a particle moving in a three-dimensional Euclidean space. Let $\dot\gamma$ be, as before, a vector tangent to the trajectory $\gamma(\tau)$ at point τ. Since, according to the special theory of relativity, no signal can propagate at velocity exceeding the speed of light c, we have $c\,dt > \sqrt{dx^2 + dy^2 + dz^2}$, where $c\,dt$, dx, dy, dz are the coordinates of an infinitesimal translation along the trajectory $\gamma(\tau)$ in the direction of the tangent vector $\dot\gamma(\tau)$. In other words, the length of any path

cannot exceed the distance travelled by light for a given time. Hence, we find that the relation $-c^2\,dt^2 + dx^2 + dy^2 + dz^2 < 0$ always holds true along the world line $\gamma(\tau)$ of any particle, i.e. $\langle\dot{\gamma}(\tau), \dot{\gamma}(\tau)\rangle_1 < 0$. This means that any vector tangent to a world line is a time-like vector and the world line of a particle always has imaginary length. In particular, a world line lies entirely inside the light (isotropic) cone with the axis t. This condition

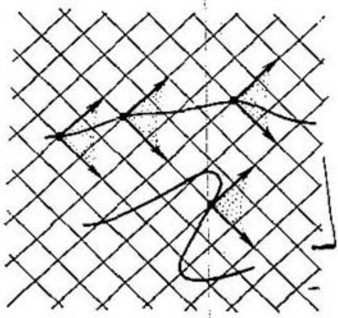

Figure 1.17. The world line depicted in the bottom cannot exist

must be satisfied at any point of a world line (see Fig. 1.17). We recall that the isotropic cone can be defined at any point of a pseudo-Euclidean space. In sections to follow we shall discuss the geometry of Minkowski's space at greater length.

Problems

1. Demonstrate that the length of a curve is the limit of the sum of the lengths of segments of a broken line connecting finitely many points on the curve when the maximal segment length tends to zero.

2. Prove that in Euclidean space a straight line is the shortest among all other lines connecting two points.

1.3. GEOMETRY ON A SPHERE AND ON A PLANE

We start with a two-dimensional Euclidean plane referred to Cartesian coordinates x, y and endowed with Euclidean metric $dl^2 = dx^2 + dy^2$ (note that the symbol ds^2 is also used sometimes

for an infinitesimal elementary arc of a smooth curve; we shall use both symbols). This Riemannian metric (apparently, it is positive definite) induces the scalar product $\langle \xi, \eta \rangle = \xi^1\eta^1 + \xi^2\eta^2$. The concept of scalar product naturally leads to the concept of circle as the set of points which are the ends of vectors of length R. If we introduce polar coordinates on a plane, then circles with centre at point O will be coordinate lines of the form $r(t) = \text{const}$. In a polar coordinate system the infinitesimal elementary arc of a circle is $rd\varphi$.

Consider now a standard embedding of a two-dimensional sphere in a three-dimensional space referred to Cartesian coordinates x, y, z. The sphere is defined as a set of points which are the ends of vectors of length R emerging from point O. Before studying the geometry of a two-dimensional sphere in greater detail, let us discuss the following general topic. Suppose a smooth curve $\gamma(t)$ lies on a sphere S^2 and we need to calculate the length of the entire curve (or of its part). Since we deal with an Euclidean space, we may consider the ambient three-dimensional Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$, write the smooth curve parametrically $\gamma(t) = (x(t), y(t), z(t))$, and calculate the quantity in question by the methods described above. The same procedure can be used if we need to measure the angle between two curves $\gamma_1(t)$ and $\gamma_2(t)$ intersecting on a sphere S^2 (both curves are assumed to lie entirely on S^2). To this end, we have to find the Cartesian coordinates of the vectors $\dot{\gamma}_1$ and $\dot{\gamma}_2$ (in a three-dimensional Euclidean space) and calculate this angle, using again the ambient Euclidean metric.

It should be noted, however, that in such calculations we use the properties of the Euclidean metric only at points in the vicinity of S^2. In other words, we could consider the Euclidean metric only for points of a two-dimensional sphere and express the metric in terms of coordinates on this sphere. Since the sphere S^2 is given in $\mathbf{R}^3$ by a single equation, the position of a point on the sphere is determined by two parameters (i.e. less by one than in three-dimensional space). This is especially clear with spherical coordinates (r, θ, φ) in $\mathbf{R}^3$. In this case a two-dimensional sphere of radius R is defined by a single equation $r = R = \text{const}$. Let us now obtain an explicit expression for the scalar product of two vectors tangent to curves lying entirely on S^2 (these vectors are therefore tangent to the sphere). Let $\gamma_1(t) = (R, \theta_1(t), \varphi_1(t))$, $\gamma_2(t) = (R, \theta_2(t), \varphi_2(t))$. We have then

$$\dot{\gamma}_1(t) = (0, \dot{\theta}_1, \dot{\varphi}_1), \quad \dot{\gamma}_2(t) = (0, \dot{\theta}_2, \dot{\varphi}_2), \text{ i.e. } \langle \dot{\gamma}_1, \dot{\gamma}_2 \rangle$$
$$= R^2 (\dot{\theta}_1\dot{\theta}_2 + \sin^2\theta(t)\,\dot{\varphi}_1\dot{\varphi}_2),$$

where $(\theta(t), \varphi(t))$ are the coordinates of the point of intersection of the curves $\gamma_1(t)$ and $\gamma_2(t)$. This implies that the scalar product just

calculated coincides with the scalar product of two vectors $(\dot{\theta}_1, \dot{\varphi}_1)$ and $(\dot{\theta}_2, \dot{\varphi}_2)$ relative to a new bilinear form: $R^2 (\dot{\theta}_1\dot{\theta}_2 + \sin^2 \theta (t)\dot{\varphi}_1\dot{\varphi}_2)$. This form induces the quadratic form $R^2 (d\theta^2 + \sin^2 \theta \, d\varphi^2)$ which is obtained from the corresponding quadratic form in Euclidean space $dr^2 + r^2 (d\theta^2 + \sin^2 \theta \, d\varphi^2)$ by substituting the functions of θ, φ: $r = R = \text{const}$, $\theta = \theta$, $\varphi = \varphi$ for the variables r, θ, φ. The Riemannian metric $R^2 (d\theta^2 + \sin^2 \theta \, d\varphi^2)$ thus obtained on S^2 is said to be an *induced* ambient Euclidean metric of a three-dimensional space.

Since the position of a point on S^2 can be specified by two parameters θ and φ (latitude and longitude), the radius vector of a point on a sphere can be represented in the form: $x = x(\theta, \varphi) = R \cos \theta \cos \varphi$, $y = y(\theta, \varphi) = R \cos \theta \sin \varphi$, $z = z(\theta, \varphi) = R \sin \theta$. Substitution of these three functions (of two parameters) into the expression for the squared elementary arc in a three-dimensional space $dx^2 + dy^2 + dz^2$ yields $(dx(\theta, \varphi))^2 + (dy(\theta, \varphi))^2 + (dz(\theta, \varphi))^2 = R^2 (d\theta^2 + \cos^2 \theta \, d\varphi^2)$.

This concrete example will be generalized below and will serve as a particular case of "induced Riemannian metrics". To this end, in the following chapters we shall introduce a rigorous concept of a "surface" to which the ambient Riemannian metric is restricted. A two-dimensional sphere standardly embedded in a three-dimensional space is precisely a "surface". We shall give some other examples of "surfaces", which are mainly defined as vector functions of several variables. For example, if the vector function (radius vector) $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ with the parameters u, v varying in a domain on a plane is given in a three-dimensional Euclidean space, this radius vector sweeps out (as u and v change) a set which we shall call (provided the vectors $\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right)$ and $\left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}\right)$ are linearly independent at each point (u, v)) a "two-dimensional surface" in a a three-dimensional space. (The general definition will be given below.) Once such a surface (or a piece of it) is defined, there arises the following quadratic form induced by the Euclidean Riemannian metric: $(dx(u, v))^2 + (dy(u, v))^2 + (dz(u, v))^2$.

Let us consider a two-dimensional sphere S^2 standardly embedded in $\mathbf{R}^3$ and endowed with an induced Riemannian metric. We have already obtained an explicit expression for this metric: $R^2 (d\theta^2 + \sin^2 \theta \, d\varphi^2)$, where θ, φ are spherical coordinates. Some other curvilinear coordinates can also be introduced on a sphere S^2. Here are several examples.

We first consider the stereographic projection of a sphere S^2 onto a plane $\mathbf{R}^2$. To do this, we place the centre of the sphere of radius R at the origin O and consider the coordinate plane $\mathbf{R}^2 (x, y)$ through

O; the north and south poles on S^2 are denoted by N and S, respectively. Let P be an arbitrary point on the sphere, distant from N. Connect the north pole N with P and extend the segment NP to the point Q where it intersects the plane $\mathbf{R}^2\,(x, y)$. Associating point Q with P, we obtain a mapping φ_0: $S^2 \to \mathbf{R}^2$ which is called the *stereographic projection* of a sphere onto a plane. The construction implies that the mapping φ_0 is defined at all points of the sphere except at the north pole N. We may assume, conditionally of course, that the north pole "maps" infinitely distant points of a two-dimensional plane (see Fig. 1.18). Introducing coordinates both on the sphere and on the plane, we can write the mapping φ_0 in analytic form.

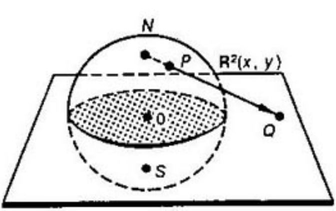

Figure 1.18

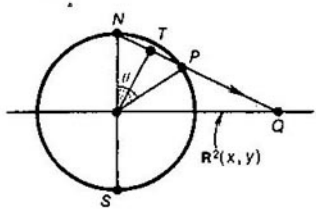

Figure 1.19

Consider, for example, spherical coordinates r, θ, φ in $\mathbf{R}^3$. These coordinates induce coordinates on the sphere S^2 and on the plane $\mathbf{R}^2\,(x, y)$. Indeed, the coordinates (θ, φ) arise on the sphere and polar coordinates (r, φ) arise on $\mathbf{R}^2$. Since the mapping φ_0 preserves the coordinate φ, it is sufficient, for determining φ_0, to find an explicit relation between the radius r and angle θ. Consider the section of S^2 by a plane through points P, O, and N (Fig. 1.19). Since the angle ONT is equal to $\pi/2 - \theta/2$, we obtain from the right-angled triangle ONQ: $r = OQ = R \tan(\pi/2 - \theta/2) = R \cot\frac{\theta}{2}$. Thus, the final equations for the coordinate transformation are: $\varphi = \varphi$, $r = R \cot(\theta/2)$.

The Jacobi matrix of the coordinate transformation is

$$\begin{pmatrix} 1 & 0 \\ 0 & -\dfrac{R}{2\sin^2\dfrac{\theta}{2}} \end{pmatrix}, \text{ i.e. } J = -\frac{R}{2\sin^2\dfrac{\theta}{2}},$$

and the transformation is regular at all points except at the north pole. Hence, we can introduce on the sphere S^2 coordinates induced by polar coordinates on a Euclidean plane. The Riemannian metric

of sphere in these coordinates takes the form

$$ds^2 = R^2 (d\theta^2 + \sin^2\theta\, d\varphi^2), \quad dr = \frac{-R}{2\sin^2\frac{\theta}{2}}\, d\theta,$$

$$\sin^2\frac{\theta}{2} = \frac{R^2}{R^2+r^2}, \quad \cos^2\frac{\theta}{2} = \frac{r^2}{R^2+r^2},$$

$$ds^2 = \frac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\, d\varphi^2).$$

Note that this Riemannian metric differs from the Euclidean metric on a plane in polar coordinates ($dr^2 + r^2\, d\varphi^2$) only by the variable factor $\frac{4R^4}{(R^2+r^2)^2}$. Such metrics are called conformal.

Definition 5. The Riemannian metric $g_{ij}(z)$ defined in a Euclidean domain C in curvilinear coordinates $z^1, \ldots, z^n$ is called *conformal* if it can be represented in the form $g_{ij}(z) = \lambda(z)\, g^{l}_{ij}(z)$, where $\lambda(z)$ is a smooth function on C and $g^{l}_{ij}(z)$ are the components of the Euclidean metric written in the coordinates $z^1, \ldots, z^n$. In other words, the metric $g_{ij}(z)$ is called conformal if there exists a coordinate system x such that $g_{ij}(x) = \lambda(x) \sum_{k=1}^{n} (dx^k)^2$.

Thus, on a Euclidean plane (referred to polar coordinates) we can consider two Riemannian metrics: $dr^2 + r^2\, d\varphi^2$ (the Euclidean metric) and $\frac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\, d\varphi^2)$ (the metric of a sphere). Both can be assumed to be defined on the same domain, $\mathbf{R}^2(x, y)$. Let us now turn back to the topic considered above, namely, the equivalence of metrics. We may ask: does a regular coordinate transformation exist on the plane $\mathbf{R}^2$ which sends the metric $dr^2 + r^2\, d\varphi^2$ into the metric $\frac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\, d\varphi^2)$? Here we may only say, intuitively, that these two metrics are not equivalent. Let us calculate the circumference of the circle $x^2 + y^2 = a^2$ in the Euclidean and "spherical" metrics. By circle we mean a smooth trajectory on a plane $\mathbf{R}^2$ the length of which can be calculated in different metrics introduced on $\mathbf{R}^2$. We shall calculate this quantity as a function of the circle radius. The Euclidean formula is quite familiar: $l_e = 2\pi a$, where a is the radius (calculated in the Euclidean metric). To find the circumference in the "spherical" metric, we first derive a relation between the Euclidean radius a and the radius ρ in the spherical metric (see Fig. 1.20)

$$\rho = 2\int_0^a \frac{R^2}{R^2+r^2}\, dr = 2R\tan^{-1}\left(\frac{a}{R}\right),$$

$$l_c = 2\int_0^{2\pi} \frac{R^2 \cdot a\, d\varphi}{R^2+a^2} = \frac{4\pi a R^2}{R^2+a^2} = 2\pi R \sin\left(\frac{\rho}{R}\right).$$

It is clear from the figure that ρ is equal to the length of the meridian connecting the north pole with a variable point of the circle. Our formal calculations in terms of the Riemannian metrics on a plane can also be performed with the aid of elementary geometric relations. Thus, if we could define a circle of radius a with centre at a certain point in terms of quantities independent of a local coordinate system, we would obtain a formula for the circumference which does not depend on the local coordinate system either. Such a definition does exist. Let us call the minimal length of a curve connecting

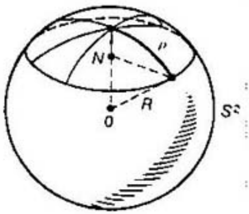

Figure 1.20

points P and Q the distance between these points (the minimum is determined for all smooth curves connecting the points). Then the circle of radius a with centre at point P is the set of points Q such that their distance from P is equal to a. To apply this definition to a sphere, we have to prove that the distance between points P and Q on the sphere S^2 is equal to the length of the arc of the large circle through these points. This proposition is easily proved if any curve connecting P and Q is approximated by a broken line composed of finitely many arcs of large circles.

In particular, for $\rho \to 0$ (i.e. for a circle of a small radius in comparison with R) we obtain $l_c \sim 2\pi\rho$, that is, the formula derived above coincides with the Euclidean expression for the circumference. Comparison of "the Euclidean circumference of radius ρ, $2\pi\rho$" with "the spherical circumference of the same radius, $2\pi R \sin \rho/R$" shows that these two functions are essentially different: one is linear and the other is periodic.

One can easily understand why a convex spherical surface cannot be deformed into a domain on a Euclidean plane in such a way that the length of a curve on the spherical surface is preserved if one recalls that it is much more difficult to flatten a spherical segment than a cylindrical segment.

The above calculation of the circumference on a sphere could also be carried out directly in spherical coordinates (θ, φ) in which the metric of a sphere is $R^2 (d\theta^2 + \sin^2 \theta \, d\varphi^2)$. This metric can be defined

on a disk of radius π on a Euclidean plane referred to the coordinates θ, φ. Obviously, in these coordinates the circumference of a circle of radius θ is equal to $2\pi R \sin\theta$ (Fig. 1.21); the point O is, naturally, identified with the north pole and the circumference of the boundary circle (of radius π) is zero, since $d\theta = 0$, $\sin(\pi) = 0$, so that the entire circle is transformed into a single point identified with the south pole of the sphere.

Let us now turn to calculating the length of a smooth curve on a sphere. By way of example, we shall consider the so-called loxodrome, the trajectory intersecting each meridian of a sphere at the

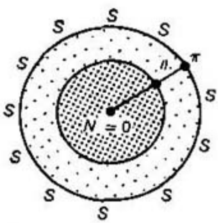

Figure 1.21

same angle α. (**Problem**. Find the length of this curve between points $\gamma(a)$ and $\gamma(b)$.) This curve is well known in navigation theory, for it is very convenient for choosing flight routes between two fixed points. The fact is that the angle α can easily be measured: it is equal to the angle between the aircraft velocity vector and the compass direction. Thus, to travel along a pre-determined loxodrome connecting the starting and terminal points, we may utilize only this parameter (the real situation is, of course, much more complicated).

A convenient method of finding an explicit parametric expression for a loxodrome is to project stereographically the sphere (and hence the loxodrome) onto a plane referred to polar coordinates (r, φ). After this operation, the loxodrome is mapped into a plane trajectory. We have already calculated the metric of a sphere in these coordinates: $\frac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\,d\varphi^2)$. The stereographic projection transforms meridians into rays emerging from the origin on the plane. We assert that the image of a loxodrome is a curve on the plane, which intersects all these rays at the same angle α (this curve is therefore uniquely determined by this angle, up to a rotation). This assertion follows from a more general fact: a stereographic projection preserves angles between intersecting curves. The transformation satisfying this condition is called conformal. To be more exact, we

mean the following. Let us consider a spherical metric (see above) induced on a sphere by the ambient Euclidean metric, and let two curves intersect at an angle $\alpha(\dot\gamma_1, \dot\gamma_2)$ (the angles are assumed oriented) (see Fig. 1.22). The number $\alpha(\dot\gamma_1, \dot\gamma_2)$ has been defined in Section 1.2. The mapping φ_0 (stereographic projection) transforms

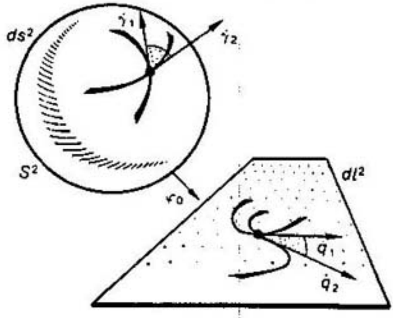

Figure 1.22

the curves γ_1 and γ_2 into curves q_1 and q_2; the angle between the latter, as calculated in the Euclidean metric dl^2 on a plane, is denoted by $\beta(\dot q_1, \dot q_2)$.

Lemma 1. *For any intersecting curves γ_1 and γ_2 the equality $\alpha(\dot\gamma_1, \dot\gamma_2) = \beta(\dot q_1, \dot q_2)$ holds true.*

Proof. It is sufficient to compare the explicit formulas for α and β and use the expression for the spherical metric (under the mapping φ_0) written in polar coordinates on $\mathbf{R}^2$. We have $ds^2(r, \varphi) = \frac{4R^4}{(R^2+r^2)^2} \times (dr^2 + r^2\, d\varphi^2)$; $dl^2(r, \varphi) = dr^2 + r^2\, d\varphi^2$. It is clear that $ds^2(r, \varphi) = \lambda^2(r)\, dl^2(r, \varphi)$, whence

$$\alpha(\dot\gamma_1, \dot\gamma_2) = \frac{\lambda^2(\dot r_1 \dot r_2 + r^2 \dot\varphi_1 \dot\varphi_2)}{\sqrt{\lambda^2(\dot r_1^2 + r^2\dot\varphi_1^2)\,\lambda^2(\dot r_2^2 + r^2\dot\varphi_2^2)}} = \beta(\dot q_1, \dot q_2).$$

The lemma is proved.

A more general statement is also valid.

Lemma 2. *Let $g_{ij}(z)$ and $q_{ij}(z)$ be two metrics defined in a Euclidean domain C in curvilinear coordinates $(z^1, \ldots, z^n)$. If the identity $g_{ij} = \lambda q_{ij}$, where $\lambda = \lambda(z)$ is a smooth function, holds true at any point $(z) \in C$, the angles between intersecting curves calculated in these metrics coincide.*

The *proof* of this lemma is completely analogous to the proof of Lemma 1.

Let us now return to the loxodrome problem. According to Lemma 1, it is sufficient to find its equation on a Euclidean plane. The condition that the angle α is preserved means that $\langle(\dot{r}, \dot{\varphi}), (1, 0)\rangle = \cos\alpha = \text{const}$, where $(\dot{r}, \dot{\varphi}) = \dot{\gamma}$ is the tangent vector to the loxodrome and $(1, 0)$ is the velocity vector of the ray ($\varphi = \text{const}$, $r = t$). Hence, $\frac{\dot{r}}{\sqrt{\dot{r}^2 + r^2\dot{\varphi}^2}} = \cos\alpha = \text{const}$ and therefore $\dot{r}^2 \sin^2\alpha = r^2\dot{\varphi}^2 \cos^2\alpha$, $\dot{r}/r = \dot{\varphi}\cot\alpha$, $(\ln r)' = \dot{\varphi}\cot\alpha$, $r = c\cdot e^{\varphi\cot\alpha}$, where $c = \text{const}$. Putting $\varphi = t$, we have $\dot{r} = c\cdot\cot\alpha\cdot e^{\varphi\cot\alpha}$, $\dot{\varphi} = 1$. Calculating the length of the arc, we obtain $l(\gamma)_{\varphi_0}^{\varphi} = c'\cdot e^{\varphi\cot\alpha} + c''$, where c'', $c' = \text{const}$. (Exercise. Find the constants c' and c''.)

Problems

1. Demonstrate that the sum of the angles of a triangle formed by arcs of large circles on a sphere S^2 exceeds 2π.
2. Express the sum of the angles of a triangle on S^2 in terms of its area (the triangle is formed by arcs of large circles).
3. Demonstrate that the similarity transformation on a sphere S^2 is possible if only the similarity ratio is equal to unity.
4. Prove that for any Riemannian metric there can be found a coordinate system such that the matrix of the Riemannian metric at a given point is an identity matrix.

1.4. PSEUDOSPHERE AND LOBACHEVSKIAN GEOMETRY

Let us consider a pseudo-Euclidean space $\mathbf{R}_s^n$ of index s. In a Euclidean space $\mathbf{R}^n$ a sphere S^{n-1} (hypersphere) can be defined as a set of points located at a distance ρ from the origin. In a pseudo-Euclidean space we can also consider a set of points at a distance ρ from the origin (here the number ρ need not necessarily be real, it may also be imaginary or zero). This set of points is called a *pseudosphere* of index s and is denoted by S_s^{n-1}. Clearly, $S_0^{n-1} = S^{n-1}$. Below we shall distinguish between pseudospheres of real, imaginary, or zero radius. A pseudosphere of zero radius is described by the second-order equation $\sum_{i=1}^{s} (x^i)^2 + \sum_{j=s+1}^{n} (x^j)^2 = 0$, where $x^1, \ldots, x^n$ are Cartesian coordinates in $\mathbf{R}^n$ on which the pseudo-Euclidean space $\mathbf{R}_s^n$ is modelled. Obviously, a pseudosphere of zero radius coincides with an isotropic (zero) cone.

Here are some examples. Let $n = 2$ and $s = 1$. The isotropic cone consists of two straight lines $x^1 = \pm x^2$ (we consider a two-dimensional plane $\mathbf{R}^2$ referred to Cartesian coordinates x^1, x^2, and model pseudo-Euclidean geometry of index one on this plane). This cone subdivides $\mathbf{R}^2$ into two domains: the domain defined by the

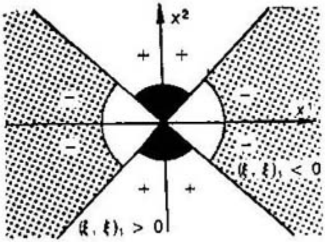

Figure 1.23

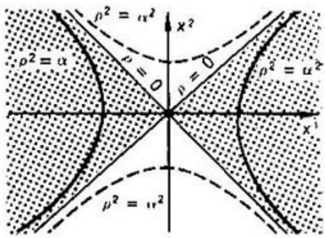

Figure 1.24

inequality $|x^2| > |x^1|$ in which $\langle\xi, \xi\rangle_1 > 0$, and the domain $|x^2| < |x^1|$ in which $\langle\xi, \xi\rangle_1 < 0$ (Fig. 1.23). Pseudospheres of real radius are hyperbolas $-(x^1)^2 + (x^2)^2 = \alpha^2$, where α is a real number, and pseudospheres of imaginary radius are hyperbolas $-(x^1)^2 + (x^2)^2 = -\alpha^2$ (Fig. 1.24).

Let $n = 3$ and $s = 1$. The isotropic cone (a pseudosphere of zero radius) is an ordinary cone of the second order given by the equation

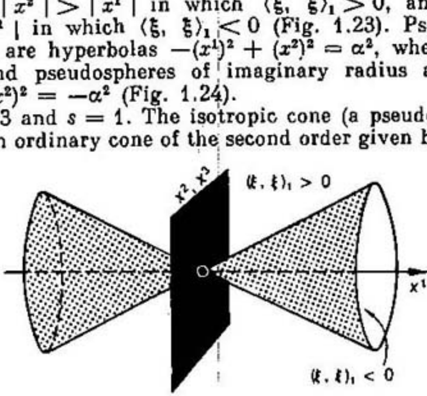

Figure 1.25

$(-(x^1)^2 + (x^2)^2 + (x^3)^2 = 0$ (x^1 is the cone axis). This cone also subdivides the entire space into two domains (using the customary terms, we may say "subdivides into the interior and exterior domains") (Fig. 1.25). Pseudospheres of real radius are one-sheet hyperboloids $-(x^1)^2 + (x^2)^2 + (x^3)^2 = +\alpha^2$, and those of imaginary radius are two-sheet hyperboloids $-(x^1)^2 + (x^2)^2 + (x^3)^2 = -\alpha^2$ $(\rho^2 =$

$-\alpha^2$) (see Fig. 1.26). Let us now discuss the metric properties of space $\mathbf{R}_1^3$, which is modelled in $\mathbf{R}^3$; if the Cartesian coordinates in $\mathbf{R}^3$ are denoted by x, y, and z, we have $\langle \xi, \xi \rangle_1 = -x^2 + y^2 + z^2$.

Consider a pseudosphere of imaginary radius. It is a two-sheet hyperboloid given by the equation $-\alpha^2 = -x^2 + y^2 + z^2$. Since this hyperboloid is embedded in $\mathbf{R}_1^3$, we may say that "the geometry of the space $\mathbf{R}_1^3$ induces a certain geometry on a pseudosphere of

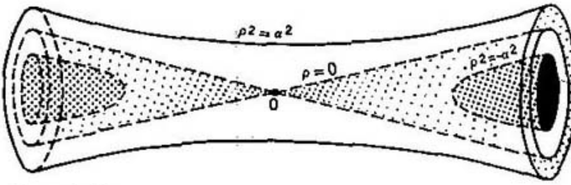

Figure 1.26

imaginary radius". This idea can also be expressed in terms of indefinite metric defined in $\mathbf{R}_1^3$: "the metric of $\mathbf{R}_1^3$ induces a metric on a pseudosphere". The phrase "geometry induced on a hyperboloid" can be given a reasonable meaning even without introducing the concept of indefinite metric. Consider a pseudosphere $-\alpha^2 = -x^2 + y^2 + z^2$ (for the sake of simplicity, we shall be dealing only with one sheet, say, the sheet defined by the inequality $x > 0$); by "points" of the geometry induced on this pseudosphere we shall mean ordinary points of the hyperboloid and by "straight lines" of the induced geometry, all possible lines of intersection of the hyperboloid and the planes $ax + by + cz = 0$ through the origin (Fig. 1.27). It turns out that the geometry introduced in this way on a pseudosphere can be studied by the methods of analytic geometry (i.e. without resorting to the concepts of indefinite metric), using a far-reaching analogy with geometry on an ordinary sphere. To this end, it is convenient to perform a transformation similar to the stereographic projection of a sphere onto a plane. The origin, point O, is taken as the centre of the pseudosphere $S_1^2 = \{-\alpha^2 = -x^2 + y^2 + z^2\}$, the point $(-\alpha, 0, 0)$ as the north pole, and the point $(\alpha, 0, 0)$ as the south pole. The plane YOZ through the centre of the pseudosphere is chosen as the projection plane (note that the restriction of the scalar product $\langle \xi, \eta \rangle_1$ to the plane YOZ is of the form $\xi^2\eta^2 + \xi^3\eta^3$, i.e. the pseudo-Euclidean scalar product $\langle \xi, \eta \rangle_1$ induces the Euclidean scalar product on YOZ). Consider now a variable point P on the right-hand sheet of the hyperboloid and connect it with the north pole N. The segment $\overrightarrow{PN}$ meets the plane YOZ at a point which is denoted by $f(P)$ and called the image of P under the stereographic

projection $f: S_1^2 \to \mathbf{R}^2$. Exactly in the same way we can define the stereographic projection of the left-hand sheet of the hyperboloid onto the same plane YOZ, using the north pole N as a projection centre. Figure 1.28 shows the section of pseudosphere by a plane through the axis OX. The image of the right-hand sheet of the hyperboloid does not cover the entire plane YOZ, but only the interior of the disk $y^2 + z^2 < \alpha^2$ of radius α. The image of the left-hand

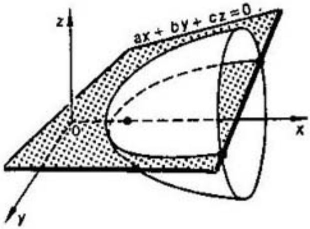

Figure 1.27

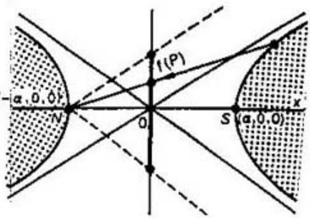

Figure 1.28

sheet of the hyperboloid covers the exterior of the circle $y^2 + z^2 = \alpha^2$. Unlike the case of an ordinary sphere S^2, the image of a pseudosphere S_1^2 under stereographic projection covers only part of the plane YOZ, since the circle $y^2 + z^2 = \alpha^2$ does not belong to the projection image. The north pole N is mapped (as in the case of a sphere) into an infinitely distant point of the plane YOZ. Let point P have Cartesian coordinates (x, y, z) $(x > 0)$ and let (u^1, u^2) be Cartesian coordinates of the point $f(P)$ on the plane YOZ, where the mapping f is a stereographic projection. The following lemma establishes an explicit relation between these coordinates.

Lemma 1. *Let* $P = (x, y, z)$, $f(P) = (u^1, u^2)$. *Then*,

$$x = \alpha \cdot \frac{|\mathbf{u}|^2 + \alpha^2}{\alpha^2 - |\mathbf{u}|^2}, \quad y = \frac{2\alpha^2 u^1}{\alpha^2 - |\mathbf{u}|^2}, \quad z = \frac{2\alpha^2 u^2}{\alpha^2 - |\mathbf{u}|^2},$$

where $|\mathbf{u}|^2 = (u^1)^2 + (u^2)^2$; $\mathbf{u} = (u^1, u^2)$.

Proof. From Fig. 1.28 it follows that

$$\frac{y}{u^1} = \frac{x+\alpha}{\alpha}, \quad \frac{z}{u^2} = \frac{x+\alpha}{\alpha}, \quad \text{i.e. } y = u^1 \frac{x+\alpha}{\alpha}, \quad z = u^2 \frac{x+\alpha}{\alpha}.$$

Since $-\alpha^2 = -x^2 + y^2 + z^2$, we have

$$-\alpha^2 = -x^2 + ((u^1)^2 + (u^2)^2) \frac{(x+\alpha)^2}{\alpha^2},$$

whence (because $(x - \alpha) > 0$) $x - \alpha = \frac{x+\alpha}{\alpha^2} \cdot |\mathbf{u}|^2$, i.e.

$$x = \alpha \cdot \frac{|\mathbf{u}|^2 + \alpha^2}{\alpha^2 - |\mathbf{u}|^2}.$$

The lemma is proved.

Lemma 2. *The coordinates (u^1, u^2) (varying in the open disk $(u^1)^2 + (u^2)^2 < \alpha^2$) define a regular coordinate system on the right-hand sheet of a hyperboloid, that is, the stereographic projection defines a regular coordinate transformation $f(x, y, z) \to (u^1, u^2)$.*

Proof. The coordinates (x, y, z) of point P on the right-hand sheet of the pseudosphere are related by $-\alpha^2 = -x^2 + y^2 + z^2$, so that the position of point P is uniquely determined by two numbers, say, y and z, i.e. the coordinates of the orthogonal projection of P onto the plane YOZ, and we may assume that the right-hand sheet of the pseudosphere is given by the equation $x = \sqrt{\alpha^2 + y^2 + z^2}$. Thus, the stereographic projection f can be meant as the coordinate transformation $(y, z) \to (u^1, u^2)$. It remains to find the Jacobi matrix of this transformation and verify that its Jacobian is not zero. Straightforward calculation yields

$$\frac{\partial y}{\partial u^1} = \frac{2\alpha^2(\alpha^2 + (u^1)^2 - (u^2)^2)}{(\alpha^2 - |\mathbf{u}|^2)^2}, \quad \frac{\partial y}{\partial u^2} = \frac{2\alpha^2 2u^1u^2}{(\alpha^2 - |\mathbf{u}|^2)^2},$$

$$\frac{\partial z}{\partial u^1} = \frac{2\alpha^2 2u^1u^2}{(\alpha^2 - |\mathbf{u}|^2)^2}, \quad \frac{\partial z}{\partial u^2} = \frac{2\alpha^2(\alpha^2 + (u^2)^2 - (u^1)^2)}{(\alpha^2 - |\mathbf{u}|^2)^2},$$

$$J(f) = 4\alpha^4 \frac{\alpha^2 + |\mathbf{u}|^2}{(\alpha^2 - |\mathbf{u}|^2)^3} = \frac{4\alpha^3 x}{(\alpha^2 - |\mathbf{u}|^2)^2} > 0.$$

Thus, the transformation Jacobian is positive at all points of the disk $(u^1)^2 + (u^2)^2 < \alpha^2$. This completes the proof of the lemma.

Before proceeding further, let us consider once more a sphere S^2. What geometry will arise on a sphere if by "points" we mean ordinary points of the sphere and by "straight lines" cross-sections of the sphere by planes through the origin O (i.e. equators)? Let us dwell on an elementary level and elucidate the geometric axioms which this set of "points" and "straight lines" satisfies. Clearly, one and only one "straight line" passes through a pair of non-antipodal "points", but infinitely many "straight lines" may pass through antipodal "points". Moreover, it is not possible to draw a "straight line" through a "point" not lying on a given "line", without intersecting the original "straight line". In other words in this geometry on a sphere "parallel (i.e. non-intersecting) straight lines" do not exist.

Some improvements are possible if this geometry could be approximated, as close as possible, to the Euclidean geometry. For example, if in the new geometry by "points" we mean pairs of antipodal points $(P, -P)$ on S^2, the first "drawback" is eliminated, that is, one and only one "straight line" passes through any two "points" (provided these "points" do not coincide). This property is analogous to the corresponding property of the Euclidean geometry. It is a simple matter to observe that in the new geometry all classical Euclidean axioms are satisfied, except the so-called "fifth postulate": namely,

given a "point" outside a "straight line", no "straight line" parallel to the original one can be drawn through this point, that is, any two "straight lines" either intersect at one "point" or coincide. Indeed, any two (non-coinciding) equators on a sphere define one and only one point $(P, -P)$. The geometry thus constructed (sometimes called *elliptic geometry*) is as rich as the Euclidean geometry, though in the former many customary properties of the Euclidean plane are replaced by other properties which are, perhaps, more "exotic" from the classical point of view formed during the development of science on the basis of simple physical ideas about the surrounding world. Everyday human experience seems to be more prone to the "Euclidean" concepts.

The above operation of identifying points P and $-P$, where P runs a sphere, is equivalent to the factorization of a sphere with respect to reflection symmetry at point O. Since any pair $(P, -P)$ defines one and only one straight line in a three-dimensional space (in which the sphere is standardly embedded), we can associate with each "straight line" of elliptic geometry (equator) the orthogonal straight line through point O. Thus, elliptic geometry can be modelled on a two-dimensional real projective space encountered in analytic geometry. Below we shall frequently deal with this type of geometry. The above operation of associating a "straight line" with a "point" becomes a duality on a projective space. This duality permits one to deduce directly from any theorem of elliptic geometry a new theorem by substituting "straight lines" for "points" and vice versa. The statement thus obtained is not, generally, equivalent to the original one.

Let us now return to pseudo-Euclidean geometry and the geometry it induces on a pseudosphere of imaginary radius. The stereographic projection $f\colon {}^{+}S_1^2 \to \{y^2 + z^2 < \alpha^2\} = D^2$ (here ${}^{+}S_1^2$ stands fo the right-hand sheet of hyperboloid) transforms points of a hyperboloid into interior points of a two-dimensional disk D^2 of radius α. What are the curves on the circle D^2 into which the "straight lines" of our geometry on a hyperboloid are transformed? By "straight lines" we mean the lines of intersection of the hyperboloid with planes through the centre O of the pseudosphere (i.e. analogues of equators on a sphere).

Lemma 3. *Each line of the intersection of ${}^{+}S_1^2$ with the plane $ax + by + cz = 0$ is mapped by f into a circular arc intersecting the circle $y^2 + z^2 = \alpha^2$ at right angles* (Fig. 1.29).

Proof. We recall that by the angle between intersecting smooth curves we mean the angle between their velocity vectors at the intersection point. Lemma 1 implies that in order to find the curve into which a "straight line" on ${}^{+}S_1^2$ is mapped, it is sufficient to substitute into the equation of the plane $ax + by + cz = 0$ explicit expressions for the variables x, y, z in terms of u^1, u^2. Suppose

$a \neq 0$. Then the equation

$$a\alpha \cdot \frac{|\mathbf{u}|^2+\alpha^2}{\alpha^2-|\mathbf{u}|^2} + \frac{2b\alpha^2 u^1}{\alpha^2-|\mathbf{u}|^2} + \frac{2c\alpha^2 u^2}{\alpha^2-|\mathbf{u}|^2} = 0$$

is reduced, after simple algebraic transformations, to

$$\left(u^1+\frac{b\alpha}{a}\right)^2+\left(u^2+\frac{c\alpha}{a}\right)^2=\frac{\alpha^2}{a^2}(b^2+c^2-a^2),$$

i.e. it defines a circle of radius $\frac{\alpha}{a}\sqrt{b^2+c^2-a^2}=r$ centred at point $\left(-\frac{b\alpha}{a}, -\frac{c\alpha}{a}\right)$ which intersects the circle $y^2+z^2=\alpha^2$ at points A and B at right angles (Fig. 1.30). That the angle at the intersection

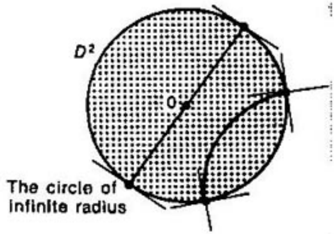

Figure 1.29

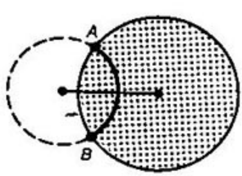

Figure 1.30

points is equal to $\pi/2$ follows from the obvious relation $\alpha^2+r^2=\frac{(b^2+c^2)\alpha^2}{a^2}$. Note that the image of a "straight line" on ${}^+S^2_1$ under the mapping f is not the entire circle $\left(u^1+\frac{b\alpha}{a}\right)^2+\left(u^2+\frac{c\alpha}{a}\right)^2=r^2$, but only its part contained in the circle $y^2+z^2<\alpha^2$.

Thus, the geometry induced on a pseudosphere of imaginary radius in $\mathbf{R}^3_1$ coincides (after an appropriate coordinate transformation) with the geometry in a circle of radius α on a Euclidean plane $\mathbf{R}^2$, provided by "points" of this geometry we mean ordinary points of this circle (except for boundary points), and by "straight lines" circular arcs intersecting the circle at right angles (in particular, all diameters of the circle are "straight lines", for these diameters can be considered as arcs of infinitely large radius). This is the so-called *Lobachevskian geometry*, and its model in a circle of radius α on a Euclidean plane is called the Poincaré model of Lobachevskian geometry. N.I. Lobachevski developed his geometry in quite another way, without using pseudo-Euclidean spaces, but proceeding from such a form of the "fifth axiom" which assumes the existence of infinitely many straight lines parallel to a given one.

Using the Poincaré model, we can easily verify that all Euclid's axioms hold true, except for the fifth axiom. Figure 1.31 clearly shows that infinitely many "straight lines" parallel to (i.e. not intersecting) a given one can be drawn through a point outside the original "straight line". From the point of view of the parallel axiom, Lobachevskian geometry is just opposite to elliptic geometry. Note also that if the parameter α tends to infinity, in any finite domain on the Poincaré model Lobachevskian geometry will "tend" to Euclidean geometry, since the arcs will be straightening, thereby turning

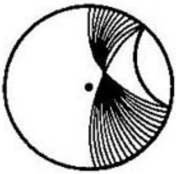

Figure 1.31

into Euclidean straight lines. The boundary of the Poincaré model, the circle $y^2 + z^2 = \alpha^2$, is called the absolute; infinitely distant points of the Lobachevskian plane are located on this boundary. While studying the Lobachevskian plane, we sometimes assume, for simplicity, $\alpha = 1$.

Remark. We could also consider geometry that arises on a pseudosphere of real radius in $\mathbf{R}_1^3$ (i.e. a one-sheet hyperboloid). (Exercise. Prove that this geometry coincides with Lobachevskian geometry.)

Let us now turn to calculating the Riemannian metric induced on a pseudosphere of imaginary radius by the ambient indefinite metric. We shall proceed by analogy with an ordinary sphere, introducing in $\mathbf{R}_1^3$ an analogue of spherical coordinates and writing with their help the equation of a pseudosphere in a convenient form. In the plane YOZ we shall introduce polar coordinates (r, φ), where φ is the angle with the axis y. Moreover, we shall introduce the parameter θ', an analogue of the corresponding parameter in ordinary spherical coordinates. Make the following transformation: $y = \alpha \sinh \theta' \cos \varphi$, $z = \alpha \sinh \theta' \sin \varphi$, $x = \alpha \cosh \theta'$. In this "pseudospherical" coordinate system the equation of a pseudosphere takes the form $\alpha = \text{const}$. This is a direct consequence of the equation of a pseudosphere of imaginary radius.

We now calculate the Riemannian metric on a pseudosphere in the coordinates u^1, u^2 in the Poincaré model. Substituting the formulas for the stereographic projection into the expression for the

squared elementary arc in $\mathbf{R}_1^3$, we obtain (verify!)

$$-(dx(u^1, u^2))^2+(dy(u^1, u^2))^2+(dz(u^1, u^2))^2=\frac{4((du^1)^2+(du^2)^2)\alpha^4}{(\alpha^2-(u^1)^2-(u^2)^2)^2}.$$

Hence, in polar coordinates (let $\alpha=1$) in the Poincaré model this metric is written as $ds^2=4\frac{dr^2+r^2\,d\varphi^2}{(1-r^2)^2}$. It can be seen from this formula that the metric is conformal, i.e. it differs from the Euclidean metric by a variable factor $\lambda(r) = 4(1-r^2)^{-2}$. To write this metric in pseudospherical coordinates, we go over from (r, φ) to new parameters (χ, φ) by $r = \tanh(\chi/2)$, $\varphi = \varphi$. Straightforward

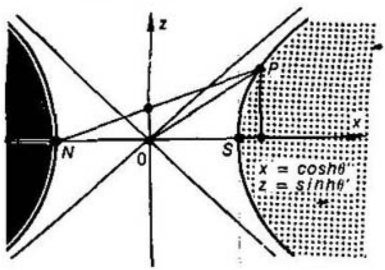

Figure 1.32

calculation yields (verify!) $ds^2 = d\chi^2 + \sinh^2\chi\, d\varphi^2$. This form of the metric is analogous to that of the metric of a sphere in the coordinates (θ, φ), but with ordinary trigonometric functions replaced by hyperbolic ones.

What is the geometric meaning of the parameter χ? In the plane XOZ the induced pseudo-Euclidean metric is of the form $ds^2 = -dx^2 + dz^2$. The section of pseudosphere by this plane is a hyperbola written parametrically in pseudospherical coordinates as (we again put $\alpha = 1$): $x = \cosh\theta'$, $z = \sinh\theta'$ (Fig. 1.32). The length of a hyperbolic segment from 0 to θ' in the pseudo-Euclidean metric is $l=\int_0^{\theta'}\sqrt{-\sinh^2\theta'+\cosh^2\theta'}\,d\theta'=\theta'$. Thus, θ' coincides with the length of a "meridian" on a pseudosphere from the south pole S to a variable point P, i.e. this parameter is completely analogous to the ordinary parameter θ on a sphere. In particular, we have elucidated the meaning of pseudo-Euclidean coordinates (see Fig. 1.33). It can be seen that pseudospherical coordinates are completely identical to spherical coordinates. If $\alpha \neq 1$, the length of a "meridian" is equal to $\alpha\theta'$.

Consider now the stereographic projection (for the sake of simplicity, we again consider only the plane XOZ because all calculations hold true if the plane XOZ is rotated about the axis OX). Since $z=\frac{2u^2\alpha^2}{\alpha^2-r^2}$, we obtain $z=\frac{2r}{1-r^2}$ (because $u^2=r$ in the plane XOZ). Furthermore, since $z=\sinh\theta'\sin\varphi=\sinh\theta'$ $\left(\text{because } \varphi=\frac{\pi}{2}\right)$, we have

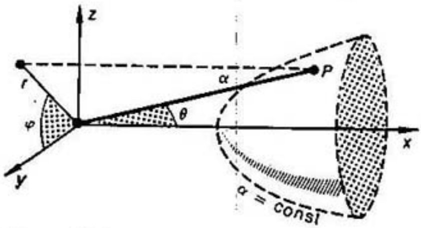

Figure 1.33

$\frac{2r}{1-r^2}=\sinh\theta'$, whence $r=\tanh\left(\frac{\theta'}{2}\right)$, so that $\chi=\theta'$. Thus, we have obtained the Riemannian metric $4\frac{dr^2+r^2\,d\varphi^2}{(1-r^2)^2}$ in the Poincaré model.

Note that this metric is positive definite, though the ambient metric is pseudo-Euclidean and, therefore, indefinite. Hence, certain surfaces (viz., a pseudosphere of imaginary radius) can carry a

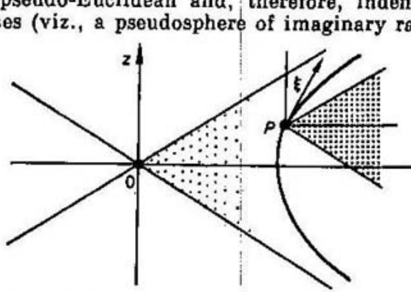

Figure 1.34

positive definite metric and be at the same time embedded in a space with indefinite metric. That the metric induced on a pseudosphere is positive definite, is clearly seen from geometric analysis. Consider, for simplicity, the section of pseudosphere by the plane XOZ, and let ξ be the velocity vector of the hyperbola at point P. Our task is to verify that the pseudo-Euclidean length of this velocity vector is real. This directly follows from Fig. 1.34 which clearly

shows that the vector ξ lies outside the light cone (with the vertex at point P) and is therefore a space-like vector.

The Riemannian metric derived above is called the Lobachevskian metric (as written in the Poincaré model). This metric can thus be considered as a new one defined on a circle, referred to ordinary polar coordinates in a Euclidean plane. We have already presented two other examples of the Riemannian metric given in a unit circle: the Euclidean metric and the spherical metric. We have proved that these metrics are not equivalent. Let us now demonstrate that the Lobachevskian metric is equivalent to neither of these two metrics. We shall use the procedure that has already been employed above: find the circumference of a circle on a Lobachevskian plane and express it in terms of the radius (calculated in the Lobachevskian metric). We assume, for simplicity, that the centre of the circle is at point O and the Euclidean radius is a. Calculate the radius in the Lobachevskian metric. By the definition of the length of a curve we have $\chi = \int_0^a \frac{2\,dr}{1-r^2} = \ln \frac{1+a}{1-a}$, i.e. $a = \tanh \frac{\chi}{2}$. The circumference $l = \int_0^{2\pi} \frac{2a\,d\varphi}{1-a^2} = \frac{4\pi a}{1-a^2} = 2\pi \sinh \chi$. If χ is rather small, we may take approximately $l \sim 2\pi\chi$, that is, we obtain the formula for the circumference in the Euclidean metric. Since, as in the case of a two-dimensional sphere, we have expressed the circumference (in the Lobachevskian metric) in terms of invariant (relative to coordinate transformation) quantities, namely, the radius in the Lobachevskian metric, the above formula for the circumference is also invariant relative to coordinate transformation, so that the Lobachevskian metric is equivalent to neither of the two metrics mentioned previously. The table of Fig. 1.35 compares the metrics of a sphere and a pseudosphere.

There are two more useful forms of the above metrics, the so-called "complex forms". Consider a Euclidean plane and introduce on it a "complex coordinate" $z = x + iy$; as $\bar{z}$ we take $x - iy$. Here we shall not concentrate on the geometric meaning of these new "coordinates", but consider the mapping $(x, y) \to (z, \bar{z})$ as a formal coordinate transformation. The Jacobi matrix of this transformation is $\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$. The Jacobian $J = -2i \neq 0$, so that the transformation is regular. Since $dz = dx + i\,dy$ and $d\bar{z} = dx - i\,dy$, the Euclidean metric in these new coordinates takes the form $ds^2 = dx^2 + dy^2 = (dx + i\,dy)(dx - i\,dy) = dz\,d\bar{z}$. Hence, the metric of a sphere is $ds^2 = 4\frac{dx^2 + dy^2}{(1 + x^2 + y^2)^2} = 4\frac{dz\,d\bar{z}}{(1 + |z|^2)^2}$, where $|z|^2 = z \cdot \bar{z} = x^2 + y^2 = r^2$.

Similarly, we obtain the complex form of the Lobachevskian metric: $ds^2 = 4\frac{dz\,d\bar{z}}{(1-|\bar{z}|^2)^2}$.

There exists another useful expression for the Lobachevskian metric on the upper half-plane. We choose another copy of a Euclidean

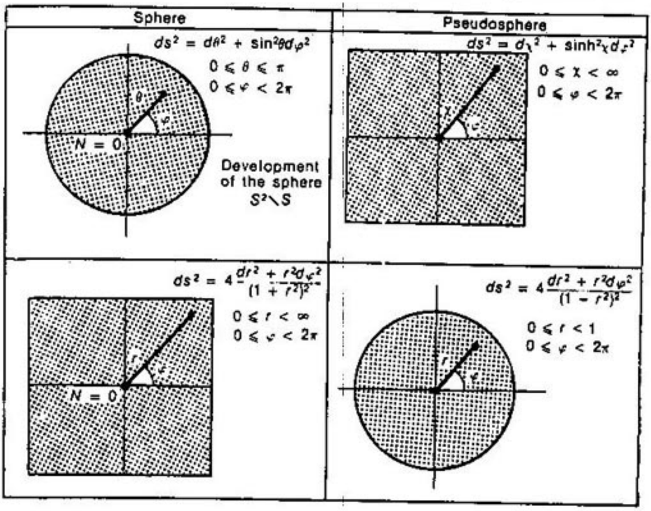

Figure 1.35

plane, introduce a new complex coordinate w, and fix the upper half-plane (i.e. the set of points such that $\operatorname{Im}(w) > 0$, where $w = u + iv$, $v = \operatorname{Im}(w)$). Consider the mapping $\mathbf{R}^2(w) \to \mathbf{R}^2(z)$ defined by $z = \frac{aw+b}{cw+d}$, where a, b, c, and d are complex numbers such that $ad - bc \neq 0$. This mapping is called homographic. If $ad - bc = 0$, the mapping transforms the entire plane $\mathbf{R}^2(w)$ into a point, that is why we have excluded this trivial case.

Our task is to find a mapping $z = \frac{aw+b}{cw+d}$ such that would transform the whole upper half-plane into the interior of the unit circle $|z| < 1$ on the plane z; in this case the real axis $\operatorname{Im}(w) = 0$ is mapped into the boundary circle $|z| = 1$. Any non-degenerate homographic mapping (i.e. the mapping for which $ad - bc \neq 0$) is uniquely defined by the image of any three points W_1, W_2, and W_3

not on the same straight line in the plane w. Here we shall not prove this general statement, since we do not need it in such a general form, but demonstrate this property for a particular example of the mapping of a half-plane onto a unit circle. Let us find the mapping $z=\frac{aw+b}{cw+d}$ such that $0\to 1$, $i\to 0$, $1\to i$ (circular permutation) (see Fig. 1.36). We obtain the following system of equations for a, b, c, and d: $1=\frac{b}{d}$, $i=\frac{a+b}{c+d}$, $0=\frac{ai+b}{ci+d}$. Solving this system (verify!), we obtain $z=\frac{1+iw}{1-iw}$. Thus, we have found a homographic mapping which transforms the upper half-plane into

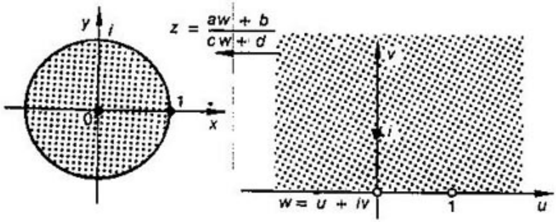

Figure 1.36

a unit circle. (There are many such mappings!) Let us prove that the mapping $z=\frac{1+iw}{1-iw}$ defines a regular coordinate transformation. Indeed, representing the mapping $z=\frac{1+iw}{1-iw}$ in the form $-1-\frac{2}{iw-1}\left(=\frac{1+iw}{1-iw}\right)$, we see that in order to prove the regularity of the coordinate transformation it is sufficient to prove that the mapping $z=\frac{1}{w}$ is regular (because the mapping in question is the following composition: $z=\frac{1}{w}$, translation by a constant vector, rotation, and extension). Writing the mapping $z=\frac{1}{w}$ in terms of the real and imaginary parts, we obtain $x=\frac{u}{u^2+v^2}$, $y=\frac{-v}{u^2+v^2}$. The Jacobi matrix is (up to a constant factor) $\begin{pmatrix} v^2-u^2 & -2uv \\ 2uv & v^2-u^2 \end{pmatrix}$, whence $J=(u^2+v^2)^2>0$. The statement is proved.

Let us find dz. We have $dz=\frac{i\,dw\,(1-iw)+i\,dw\,(1+iw)}{(1-iw)^2}=\frac{2i\,dw}{(1-iw)^2}$. Thus, $dz\,d\bar{z}=\frac{4\,dw\,d\bar{w}}{|1-iw|^4}$. Making an appropriate transformation in

the expression $ds^2 = \frac{4\,dz\,d\bar{z}}{(1-|z|^2)^2}$, we obtain $ds^2 = \frac{4\,dw\,d\bar{w}}{(wi - i\bar{w})^2} = \frac{du^2+dv^2}{v^2}$. This confirms once more that points of the real axis (the image of the absolute of the Poincaré model) are infinitely distant points on the Lobachevskian plane (modelled in this case on the upper half-plane). Indeed, the length of a segment of the axis Ov from point i to point O (not belonging to the Lobachevskain plane) is

$$l = \int_0^1 \frac{dv}{v} = \ln v \Big|_0^1 = -\ln(0) \to \infty.$$

What are the curves that we obtain when "straight lines" of a Lobachevskian plane (in the Poincaré model) are mapped onto the

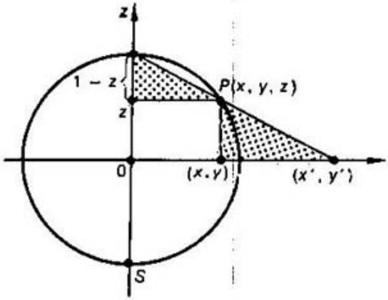

Figure 1.37

upper half-plane? First, we shall answer this question for elliptic geometry. Recall that by "straight lines" of this geometry we mean different equators on a sphere S^2. Consider the stereographic projection of S^2 onto a plane $\mathbf{R}^2$. What are the curves into which the equators are transformed?

Lemma 4. *Stereographic projection maps an equator either into a circle on* $\mathbf{R}^2$ *or into a straight line, an equator being transformed into a straight line if and only if the equator passes through the north pole of the sphere.*

Proof. We first write the stereographic projection in Cartesian coordinates. Suppose x, y, z are the coordinates of a point P on a sphere and x', y' are the coordinates of its image (after stereographic projection). From Fig. 1.37 we obtain the relations $x' = \frac{x}{1-z}$, $y' = \frac{y}{1-z}$. The equation of a circle on a sphere of radius 1 can be

written in the form $\{ax + by + cz = d,\ x^2 + y^2 + z^2 = 1\}$, whence

$$z = \frac{d - ax - by}{c}, \quad z = \frac{d - ax' - by'}{c - ax' - by'}, \quad 1 - z = \frac{c - d}{c - ax' - by'},$$

$$(x')^2 (1 - z)^2 + (y')^2 (1 - z)^2 + z^2 = 1.$$

Since $1 - z > 0$, we have $(x')^2 + (y')^2 = \frac{1+z}{1-z} = 1 + \frac{2z}{1-z} = 1 + \frac{2(d - ax' - by')}{c - d}$, i.e. $(x')^2 + (y')^2 + \frac{2a}{c-d}x' + \frac{2b}{c-d}y' = 1 + \frac{2d}{c-d}$. As is known from analytic geometry, this second-order equation defines either a straight line or a circle on a plane, depending on the relationship between the parameters a, b, c, and d. The lemma is proved.

A similar statement holds true for a Lobachevskian plane.

Lemma 5. *A non-degenerate homographic mapping* $z = \frac{aw + b}{cw + d}$ *(i.e. a mapping for which* $ad - bc \neq 0$, *where* a, b, c, *and* d *are complex numbers) transforming a two-dimensional plane into itself maps straight lines and circles into straight lines and circles; moreover, a straight line can be transformed into a circle and vice versa.*

Proof. For $c = 0$ the statement is obvious because the mapping $z = \frac{a}{d}w + \frac{b}{d}$ is a translation by the vector b/d and multiplication by the complex number a/d (extension with rotation). Let $c \neq 0$. Then, $z = \frac{a}{c} - \frac{(ad - bc)}{c(cw + d)}$, so that it remains to prove the lemma for the homographic mapping $z = \frac{1}{w}$. Consider an arbitrary circle on the plane z and write its equation in the form $|z - z_0|^2 = \varepsilon^2$, i.e. $(z - z_0)(\bar{z} - \bar{z}_0) = \varepsilon^2$. By virtue of the transformation $z = 1/w$ we obtain $(1 - z_0 z)(1 - \bar{z}\bar{z}_0) = \varepsilon^2 \cdot z\bar{z}$, whence $z\bar{z}(\varepsilon^2 - z_0\bar{z}_0) + z_0 z + \bar{z}_0\bar{z} - 1 = 0$. Clearly, this equation defines (depending on the choice of the parameters ε_0 and z_0) either a circle or a straight line. The lemma is proved.

Corollary 1. *Consider the mapping* $z = \frac{1 + iw}{1 - iw}$ *which transforms the upper half-plane into a unit circle. This coordinate transformation maps "straight lines" of the Poincaré model, i.e. circular arcs orthogonal to the absolute, either into straight lines on the plane* W, *which are orthogonal to the real axis* u $(w = u + iv)$ *or into semi-circles perpendicular to the real axis* u (see Fig. 1.38).

Proof. Suppose the unit circle is embedded in the complex plane z and the upper half-plane represents the complex variable w. The mapping $z = \frac{1 + iw}{1 - iw}$ is then defined everywhere on the plane w. According to Lemma 5, this mapping sends straight lines either into

straight lines or into circles (similarly, circles are transformed either into straight lines or circles). Since the mapping $z = \frac{1+iw}{1-iw}$ has the inverse one $w = \frac{z-1}{i(z+1)}$ (we recall that $ad - bc \neq 0$), the same statement holds true for the inverse mapping $w = \frac{z-1}{i(z+1)}$. Hence, "straight lines" of the Poincaré model (when completed to straight lines and circles on the plane z) are transformed either into straight

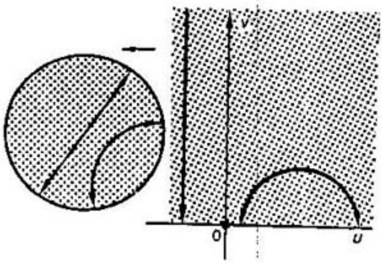

Figure 1.38

lines or circles on the plane w. It remains to prove that these "straight lines" must be orthogonal to the real axis (at points of intersection with this axis). This is a consequence of the following lemma.

Lemma 6. *Any non-degenerate homographic mapping* $z = \frac{aw+b}{cw+d}$ *which transforms the plane* w *onto the plane* z *preserves angles between smooth curves at intersection points.*

Proof. It is sufficient to calculate explicitly the Euclidean metric $dz\,d\bar{z}$ in the new coordinates $w = u + iv$. The fact that the mapping $z = \frac{aw+b}{cw+d}$, where $ad - bc \neq 0$, defines a regular coordinate system can be verified exactly in the same way as the above statement about the regularity of the mapping $z = \frac{1+iw}{1-iw}$. Straightforward calculation yields

$$dz = \frac{a\,dw\,(cw+d) - c\,dw\,(aw+b)}{(cw+d)^2} = \frac{ad-bc}{(cw+d)^2}\,dw,$$

i.e. $dz\,d\bar{z} = \frac{|ad-bc|^2}{|cw+d|^4}\,dw\,d\bar{w}$. In terms of the real coordinates (x, y) and (u, v) we have $dx^2 + dy^2 = \frac{|ad-bc|^2}{|cw+d|^2}(du^2 + dv^2)$. Thus, our mapping is conformal, i.e. it multiplies the Euclidean metric by a positive (variable) factor and, as was proved earlier, preserves

angles between intersecting smooth curves. This proves the lemma and, therefore, Corollary 1.

Consider the properties of the mapping $w=\frac{z-1}{i(z+1)}$ in greater detail. Point -1 goes over to infinity, hence, the diameter through points 0 and -1 in the Poincaré model is transformed into a

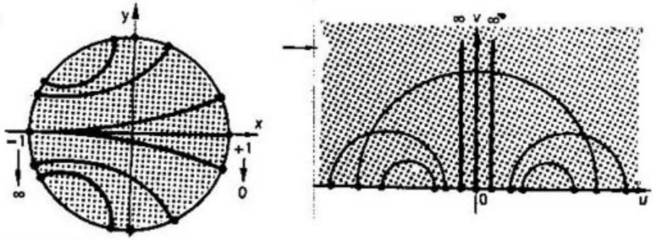

Figure 1.39

straight line orthogonal to the real axis u at point 0 (see Fig. 1.39). The mapping $w=\frac{z-1}{i(z+1)}$ sends the circle $|z|=1$ into the real axis u. Indeed, if $\lambda=\frac{e^{i\varphi}-1}{i(e^{i\varphi}+1)}$, then $\overline{\lambda}=\frac{e^{i\varphi}-1}{i(e^{i\varphi}+1)}=\lambda$, i.e. λ is real.

Let us consider a Lobachevskian plane modelled on the upper half-plane. If P_1 and P_2 are arbitrary points of the upper half-plane,

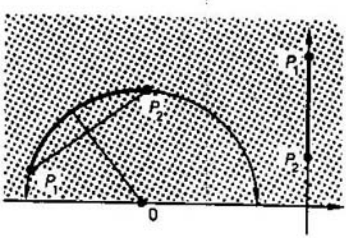

Figure 1.40

they can always be connected by a single "straight line" in the Lobachevskian plane. Figure 1.40 illustrates how to construct such a "straight line". If the two points lie on a straight line orthogonal to the real axis, a "straight line" of the Lobachevskian plane coincides with the former straight line. Let us find an explicit formula for the

length of a "straight segment" from P_1 to P_2 (in the Lobachevskian metric). We first make a simple transformation of the "straight line" which does not alter the length of its arcs, that is, translation along the axis u. Since the Lobachevskian metric is of the form $ds^2 = \frac{dx^2+dy^2}{y^2}$, it is invariant (preserved) relative to translations (in other words, it is "translation-invariant" relative to real displacements, i.e. displacements along the axis x). Suppose P_1 and P_2 do not lie on a straight line orthogonal to the real axis. In particular, we may assume that the centre of the circle representing the "straight line" through P_1 and P_2 is at point 0. The parametric equation of this "straight line" is $r = r_0$, where (r, φ) are polar coordinates. Let P_1 and P_2 have the coordinates (r_0, φ_1) and (r_0, φ_2), respectively. Then the length of the arc from P_1 to P_2 is

$$l(P_1, P_2) = \int_{\varphi_1}^{\varphi_2} \frac{r_0\,d\varphi}{r_0 \sin\varphi} = \int_{\varphi_1}^{\varphi_2} \frac{d\varphi}{\sin\varphi} = \frac{1}{2}\ln\left(\frac{1+t}{1-t}\right)\Big|_{\cos\varphi_1}^{\cos\varphi_2}$$

$$= -\frac{1}{2}\ln\frac{(1+\cos\varphi_2)(1-\cos\varphi_1)}{(1-\cos\varphi_2)(1+\cos\varphi_1)} = -\ln\left(\frac{\tan\frac{\varphi_1}{2}}{\tan\frac{\varphi_2}{2}}\right).$$

Suppose now that points P_1 and P_2 lie on a straight line orthogonal to the real axis. Obviously, $l(P_1, P_2) = \ln\left(\frac{y_2}{y_1}\right)$, where (x, y_1) and (x, y_2) are the coordinates of points P_1 and P_2, respectively. Note that if P_1 and P_2 lie on a circle with the centre on the real axis, the distance between these points remains unchanged under similarity transformation with the centre at point 0 (we assume that the centre of the circle coincides with the origin).

Using the formula for the length of an arc derived above, we can prove that for any triangle on a Lobachevskian plane formed by segments of "straight lines" the triangle inequality holds true: namely, $l(P_1, P_2) + l(P_2, P_3) \geqslant l(P_1, P_3)$, where P_1, P_2, and P_3 are the triangle vertices.

Let us prove this inequality. It is convenient, as before, to consider a Lobachevskian plane as a two-sheet hyperboloid ${}^+S_1^2$ defined by the equation $-(x^1)^2 + (x^2)^2 + (x^3)^2 = -1$. Recall that "straight lines" on a Lobachevskian plane are intersections of planes through the origin with a two-sheet hyperboloid. Let $e_1, e_2 \in {}^+S_1^2$ be two vectors with their ends on a Lobachevskian plane, and let P be a plane generated by these vectors. Then the scalar product in $\mathbf{R}_1^3$ induces an indefinite scalar product on P. Since $\langle e_1, e_2\rangle = -1$, the plane P is isomorphic to $\mathbf{R}_1^2$, so that we can introduce the linear coordinates (x, y) on P such that the metric takes the form $-x^2 + y^2$ and the equation of a "straight line" Γ is written as $-x^2 + y^2 = -1$. Hence,

the length of a segment of the curve Γ between points $\mathbf{e}_1 = (x_1, y_1)$ and $\mathbf{e}_2 = (x_2, y_2)$ can be calculated by formulas derived above. To this end, we put $x = \cosh\chi$, $y = \sinh\chi$, whence $x_1 = \cosh\chi_1$, $y_1 = \sinh\chi_1$, $x_2 = \cosh\chi_2$, $y_2 = \sinh\chi_2$, and the length of the segment is

$$l_{12} = |\chi_1 - \chi_2| = |\cosh^{-1} x_2 - \cosh^{-1} x_1|$$
$$= |\cosh^{-1}(x_1x_2 - y_1y_2) = \cosh^{-1}(-\langle \vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2\rangle).$$

Thus, for the three vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in {}^+S_1^2$ the triangle inequality becomes $l_{12} \leqslant l_{13} + l_{32}$. Substitution of the lengths in terms of scalar products of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ yields:

$$\cosh^{-1}(-\langle \mathbf{e}_1, \mathbf{e}_2\rangle) \leqslant \cosh^{-1}(-\langle \mathbf{e}_2, \mathbf{e}_3\rangle) + \cosh^{-1}(-\langle \mathbf{e}_1, \mathbf{e}_3\rangle).$$

Note that all the three scalar products $-\langle \mathbf{e}_1, \mathbf{e}_2\rangle$, $-\langle \mathbf{e}_2, \mathbf{e}_3\rangle$, and $-\langle \mathbf{e}_1, \mathbf{e}_3\rangle$ are negative on ${}^+S_1^2$. Calculating the function cosh of both the left-hand and right-hand sides of the inequality we obtain

$$-\langle \mathbf{e}_1, \mathbf{e}_2\rangle \leqslant \langle \mathbf{e}_2, \mathbf{e}_3\rangle\langle \mathbf{e}_1, \mathbf{e}_3\rangle + \sqrt{\langle \mathbf{e}_2, \mathbf{e}_3\rangle^2 - 1}\,\sqrt{\langle \mathbf{e}_1, \mathbf{e}_3\rangle^2 - 1}$$

or

$$-\langle \mathbf{e}_1, \mathbf{e}_2\rangle - \langle \mathbf{e}_2, \mathbf{e}_3\rangle\langle \mathbf{e}_1, \mathbf{e}_3\rangle \leqslant \sqrt{\langle \mathbf{e}_2, \mathbf{e}_3\rangle^2 - 1}\,\sqrt{\langle \mathbf{e}_1, \mathbf{e}_3\rangle^2 - 1}.$$

Instead of the linear inequality, it is sufficient to prove the relation

$$\langle \mathbf{e}_1, \mathbf{e}_2\rangle^2 + \langle \mathbf{e}_2, \mathbf{e}_3\rangle^2\langle \mathbf{e}_1, \mathbf{e}_3\rangle^2 + 2\langle \mathbf{e}_1, \mathbf{e}_2\rangle\langle \mathbf{e}_2, \mathbf{e}_3\rangle\langle \mathbf{e}_1, \mathbf{e}_3\rangle$$
$$\leqslant \langle \mathbf{e}_1, \mathbf{e}_3\rangle^2\langle \mathbf{e}_2, \mathbf{e}_3\rangle^2 + 1 - \langle \mathbf{e}_1, \mathbf{e}_3\rangle^2 - \langle \mathbf{e}_2, \mathbf{e}_3\rangle^2$$

or

$$-1 + 2\langle \mathbf{e}_1, \mathbf{e}_2\rangle\langle \mathbf{e}_2, \mathbf{e}_3\rangle\langle \mathbf{e}_1, \mathbf{e}_3\rangle + \langle \mathbf{e}_1, \mathbf{e}_2\rangle^2 + \langle \mathbf{e}_1, \mathbf{e}_3\rangle^2 + \langle \mathbf{e}_2, \mathbf{e}_3\rangle^2 \leqslant 0.$$

Since $\langle \mathbf{e}_1, \mathbf{e}_1\rangle = -1$, $\langle \mathbf{e}_2, \mathbf{e}_2\rangle = -1$ and $\langle \mathbf{e}_3, \mathbf{e}_3\rangle = -1$, the left-hand side of the inequality to be proved is the Gram determinant composed of scalar products of the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$:

$$\det\begin{pmatrix} \langle \mathbf{e}_1, \mathbf{e}_1\rangle & \langle \mathbf{e}_2, \mathbf{e}_1\rangle & \langle \mathbf{e}_3, \mathbf{e}_1\rangle \\ \langle \mathbf{e}_1, \mathbf{e}_2\rangle & \langle \mathbf{e}_2, \mathbf{e}_2\rangle & \langle \mathbf{e}_2, \mathbf{e}_2\rangle \\ \langle \mathbf{e}_1, \mathbf{e}_3\rangle & \langle \mathbf{e}_2, \mathbf{e}_3\rangle & \langle \mathbf{e}_3, \mathbf{e}_3\rangle \end{pmatrix} \leqslant 0.$$

The sign of the Gram determinant remains unchanged under basis transformations, so that in the standard basis it is of the form

$$\det\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = -1 \neq 0.$$

Thus, the triangle inequality is proved.

Using similar considerations for a sphere, we can prove the following statement.

Lemma 7. *The shortest distance between any two points P and Q is the segment of the "straight line" connecting these points.*

Let us consider a Lobachevskian plane and its model on the right-hand sheet of a two-sheet hyperboloid ${}^{+}S_1^2$. The "straight line" connecting P and Q is the arc of the section of ${}^{+}S_1^2$ by a plane through point O (the pseudosphere centre). Consider also an arbitrary smooth curve γ connecting points P and Q on the hyperboloid and approximate this curve by a broken line composed of "straight" segments.

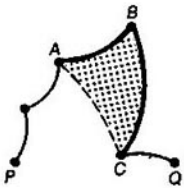

Figure 1.41

It is sufficient to prove that the length of any such broken line is not less than the length of the "straight line" connecting P and Q. Any two points on a Lobachevskian plane can be connected by a "straight" segment. We have already proved this statement in the model on the upper half-plane when we constructed, using an arbitrary pair of points on the upper half-plane, the circle passing through these points and orthogonal to the real axis (at the intersection points). We now can replace our broken line by a new broken line such that its length does not exceed the length of the original one. Consider two adjacent links of the original broken line and (if these links do not lie on a single "straight line") construct a new segment which connects the beginning of the first link with the end of the second link (Fig. 1.41). Since triangle ABC is formed by three "straight" segments, it obeys the triangle inequality (see above), so that the length of the new broken line does not exceed the length of the original one. Continuing this process and bearing in mind that the original broken line has a finite number of links, we obtain (after a finite number of steps) a broken line consisting only of one link, i.e. we have connected points P and Q by a "straight" segment. However, we have proved above that a "straight" line connecting any two points is unique. Hence, the process of "simplifying" the broken line ceased at the step where the broken line coincided with the "straight" segment connecting P and Q. The lemma is proved.

Thus, we have demonstrated why plane sections of a hyperboloid act as straight lines in Lobachevskian geometry. It turns out that precisely these lines realize the shortest distance between points on a

Lobachevskian plane. We have got accustomed to the fact that on a Euclidean plane the motion along a straight line means the motion along a trajectory which is the shortest distance between two points. Since in Lobachevskian geometry (as well as in geometry on a sphere) the concept of "rectilinear motion" cannot be formulated in a simple way in terms of vectors with constant components (which uniquely define the direction of motion), it would be quite natural to call motion along a smooth curve, which is the shortest distance between the starting and terminal points, "rectilinear motion". We have found a class of such curves both on a sphere and on a Lobachevskian plane. In both cases the class comprises plane sections of two-dimensional surfaces modelling the corresponding geometries (of sphere and of pseudosphere). In the sequel, we shall deal with the elementary calculus of variations and prove that this class of curves can be distinguished from all other smooth curves also by the property that translation (a generalization of ordinary translation in a Euclidean space) can easily be effected along such "straight lines" (called geodesics).

Problems

1. Demonstrate that the sum of the angles of a triangle on a pseudosphere composed of "straight" segments is less than 2π.

2. Express the sum of the angles of a triangle on a pseudosphere (formed by "straight" segments) in terms of its area.

Chapter 2

General Topology

Topology is a branch of mathematics that studies the properties of geometric objects which remain unaltered under "deformations" or transformations similar to deformations. There are many concepts in mathematical analysis which are analogous in their properties and investigation methods. For example, convergence and limit in analysis as (a) limit of a sequence, (b) various types of limits of a function of one variable, (c) limit of a function of several variables, (d) limit of a vector-valued function, (e) convergence of integral sums. All these concepts based on some common methods of investigation we understand intuitively as the closeness of points of a set. Another important example is various types of continuity which are very close to convergence. General topology studies the most fundamental properties of geometric spaces and their transformations related to convergence and continuity.

2.1. DEFINITION AND BASIC PROPERTIES OF METRIC AND TOPOLOGICAL SPACES

2.1.1. METRIC SPACES

"Closeness" of elements of a set is one of the fundamental concepts in topology. It can be measured most conveniently as distance between the elements.

Consider a set X of arbitrary elements. A *metric* on X is a numerical non-negative function $\rho(x, y)$ dependent on a pair of elements $x, y \in X$ such that the following conditions are satisfied:

(a) $\rho(x, y) = 0$ if and only if $x = y$,

(b) $\rho(x, y) = \rho(y, x)$,

(c) for any three elements $x, y, z \in X$ we always have $\rho(x, y) \leqslant \rho(x, z) + \rho(z, y)$.

Properties (a), (b), and (c) are called identity, symmetry, and triangle axioms, respectively.

The set X equipped with a metric ρ is called a *metric space*; the elements of X are called *points*. The value of the metric $\rho(x, y)$ on a pair of points $x, y \in X$ is called the distance between the points x

and y. A metric space should, in general, be denoted as a pair (X, ρ). But if there is no confusion, we shall frequently denote a metric space just by X.

A simple example of a metric space is the set of real numbers $\mathbf{R}^1$. The metric on this set is defined by $\rho(x, y) = |x - y|$. The function ρ possesses all the properties of metric and transforms $\mathbf{R}^1$ into a metric space.

If X is a metric space with metric ρ and $Y \subset X$ is a subset of X, then Y is also a metric space with the same metric. The metric space (Y, ρ) is called a *subspace* of metric space X. Let $Y \subset X$ be an arbitrary subset of a metric space. Consider the upper bound of all numbers $\rho(x, y)$ with x and y running points of Y. If this upper bound is finite $d = \sup\limits_{x, y \in Y} \{\rho(x, y)\}$, the set Y is called *bounded* and the number d is called the *diameter* of Y.

A *ball neighbourhood of radius* ε *with centre* $x \in X$ is the set $O_\varepsilon(x)$ of all points $y \in X$ such that $\rho(x, y) < \varepsilon$.

The *distance between two sets* $Y_1, Y_2 \in X$ is the number $\rho(Y_1, Y_2) = \inf \rho(x, y)$, where x runs Y_1 and y runs Y_2. If Y_1 and Y_2 have a common point, then $\rho(Y_1, Y_2) = 0$. There exist, however, disjoint sets with zero distance between them.

A *ball neighbourhood of a set* Y *of radius* $\vec{\varepsilon}$ is the set $O_\varepsilon(Y)$ of all points $x \in X$ such that $\rho(x, Y) < \varepsilon$.

The *point of contact* of a set Y is any point x such that $\rho(x, Y) = 0$. In particular, every point x of the set Y is the point of contact of this set. The converse statement is, in general, incorrect. For example, the interval $(a, b) \subset \mathbf{R}^1$ has points of contact a and b which do not belong to the interval (a, b).

The *closure* of a set Y is the set of all its points of contact. The closure of a set Y is denoted by $\bar{Y}$.

Thus, $Y \subset \bar{Y}$, but the converse statement is, in general, incorrect. In the above example the closure of the interval (a, b) is the segment $[a, b]$.

A set Y of a metric space is called *closed* if its closure $\bar{Y}$ coincides with Y, $Y = \bar{Y}$.

A point x is called an *interior point* of a set Y if a ball neighbourhood of x is contained in Y, i.e. $O_\varepsilon(x) \subset Y$. The set of all interior points of a set Y is called its *interior* and is denoted by $\operatorname{Int} Y$. If a set Y coincides with its interior, $\operatorname{Int} Y$, then the set is called *open*.

Theorem 1. *Let X be a metric space. The set $Y \subset X$ is closed if and only if the complement $X \setminus Y$ is open.*

Proof. Let Y be a closed set, $Y = \bar{Y}$, and $x \in X \setminus Y$. This means that x is not a point of contact of Y, i.e. $\rho(x, Y) = \varepsilon > 0$. Demonstrate that $O_\varepsilon(x) \subset X \setminus Y$. Indeed, if $y \in O_\varepsilon(x)$, then $\rho(x, y) < \varepsilon$. If $y \in Y$, then $\rho(x, y) \geqslant \rho(x, Y)$, i.e. $\rho(x, y) \geqslant \varepsilon$, which contra-

dicts the hypothesis. Hence, the set $X \setminus Y$ is open.

Conversely, let $X \setminus Y$ be an open set. Then if $x \in X \setminus Y$, there exists a ball neighbourhood $O_\varepsilon(x)$ contained in $X \setminus Y$. This means that $\rho(x, y) \geqslant \varepsilon$ for $y \in Y$, i.e. $\rho(x, Y) \geqslant \varepsilon$. Hence, x is not the point of contact of Y. Thus, if $x \in \overline{Y}$, then $x \overline{\in} (X \setminus Y)$, i.e. $x \in Y$. This means that $\overline{Y} \subset Y$, i.e. $Y = \overline{Y}$, so that the set Y is closed.

Theorem 2. *Let X be a metric space. Then the union of any family of open sets is an open set and the intersection of finitely many open sets is also an open set. In dual form: the intersection of any family of closed sets and the union of finitely many closed sets are closed sets.*

Proof. Let $Y_\alpha \subset X$ be a family of open sets. We now demonstrate that the union $Y = \bigcup_\alpha Y_\alpha$ is an open set. Suppose $x \in Y$, then $x \in Y_\alpha$ for a certain index α and since Y_α is open, the ball neighbourhood $O_\varepsilon(x)$ is entirely contained in Y, i.e. $O_\alpha(x) \subset Y_\alpha$. Therefore, $O_\varepsilon(x) \subset Y_\alpha \subset \bigcup_\alpha Y_\alpha = Y$, that is, Y is an open set. Let the index α assume finitely many values. We now verify that the intersection $Y' = \bigcap_\alpha Y_\alpha$ is an open set. If $x \in Y'$, then $x \in Y_\alpha$ for each value of α. Since Y_α is an open set, the ball neighbourhood $O_{\varepsilon_\alpha}(x)$ lies in Y_α. Take $\varepsilon = \min \varepsilon_\alpha > 0$ (here we use the fact that the set of indices α is finite). Then $O_\varepsilon(x) \subset O_{\varepsilon_\alpha}(x) \subset Y_\alpha$ for any α, i.e. $O_\varepsilon(x) \subset \bigcap_\alpha Y_\alpha = Y$. Hence, Y is an open set.

The dual statements for closed sets follow from Theorem 1 and the union, intersection, and complement are related as

$$\bigcap_\alpha (X \setminus Y_\alpha) = X \setminus \bigcup_\alpha Y_\alpha, \quad \bigcup_\alpha (X \setminus Y_\alpha) = X \setminus \bigcap_\alpha Y_\alpha.$$

In the previous statements we have not used property (c) of the metric, which is often called the triangle inequality. In many propositions to follow concerning metric spaces only the properties formulated in Theorems 1 and 2 will be used. We can, therefore, extend the class of spaces for which the concepts of open and closed sets, interior points, and points of contact are applicable. Before proceeding further, however, we prove several useful propositions about metric spaces based on the triangle inequality.

Theorem 3. *Let X be a metric space. Then,*

(a) *the ball neighbourhood $O_\varepsilon(x)$ is an open set,*

(b) *the interior* Int Y *of an arbitrary set Y is an open set,*

(c) *the closure $\overline{Y}$ of an arbitrary set Y is a closed set.*

Proof. Let $y \in O_\varepsilon(x)$, then $\rho(x, y) = \delta < \varepsilon$. Demonstrate that there exists a number $\varepsilon' > 0$ such that $O_{\varepsilon'}(y) \subset O_\varepsilon(x)$. Choose ε' such that $\delta + \varepsilon' < \varepsilon$. If $z \in O_{\varepsilon'}(y)$, then $\rho(z, y) < \varepsilon'$ and hence $\rho(z, x) \leqslant \rho(x, y) + \rho(y, z) < \delta + \varepsilon' < \varepsilon$, i.e. $z \in O_\varepsilon(x)$. This means that $O_{\varepsilon'}(y) \subset O_\varepsilon(x)$. Thus, $O_\varepsilon(x)$ is an open set.

We now prove that $\operatorname{Int} Y$ is an open set. Note that if $Y_1 \subset Y_2$, then $\operatorname{Int} Y_1 \subset \operatorname{Int} Y_2$. Hence, if $x \in \operatorname{Int} Y$, we have $O_\varepsilon(x) \subset Y$, and by virtue of statement (a) $O_\varepsilon(x) = \operatorname{Int} O_\varepsilon(x) \subset \operatorname{Int} Y$. Thus, by definition, $\operatorname{Int} Y$ is open.

Let us finally verify that $\overline{Y}$ is a closed set. By definition, x does not belong to $\overline{Y}$ if and only if $x \in \operatorname{Int}(X \setminus Y)$, i.e. $X \setminus \overline{Y} = \operatorname{Int}(X \setminus Y)$. According to Theorem 1, if $\operatorname{Int}(X \setminus Y)$ is open, then $\overline{Y}$ is closed. This completes the proof of Theorem 3.

2.1.2. TOPOLOGICAL SPACES

A *topology* is said to be defined on a set X if a family of subsets of X, called *open* sets, is valid and the following conditions are satisfied:

(a) *The entire set X and the empty set are open sets.*

(b) *The union of any family of open sets and the intersection of finitely many open sets are open sets.*

A set X with a topology defined on it is called a *topological space*, and the elements of X are called its *points*. The complement to an open set is called a *closed set*. Obviously, in any topological space X the following duality properties for closed sets hold true:

(a′) *The entire set X and the empty set are closed sets.*

(b′) *The intersection of any family of closed sets and the union of finitely many closed sets are closed sets.*

Thus, to define a topology on a set X, it is sufficient to define a family of sets satisfying conditions (a′) and (b′), instead of a family of open sets, and call the former closed sets.

Theorem 2 of the preceding section implies that a family of open sets of a metric space X induces a topology on X, that is, transforms X into a topological space.

Many important properties of a metric space are also valid in a topological space. Any open set containing x is called a *neighbourhood of point x* of a topological space X. Similarly, an open set containing a subset of Y is called a *neighbourhood of set Y*. A *point of contact* of a set $Y \subset X$ is a point x such that each its neighbourhood has a non-empty intersection with Y. The set of all points of contact of a set Y is said to be the *closure* of Y and is denoted by $\overline{Y}$. An *interior point* of a set Y is a point $x \in Y$ contained in Y together with its neighbourhood. The set of all interior points of a set Y is called the *interior* of Y and is denoted by $\operatorname{Int} Y$.

Theorem 4. *The set $Y \subset X$ is closed (i.e. it is the complement to an open set) if and only if $Y = \overline{Y}$.*

Proof. Let Y be a closed set, i.e. $X \setminus Y$ is open. Then the set $X \setminus Y$ is a neighbourhood for any its point, i.e. the points of $X \setminus Y$ are not points of contact of the set Y. Therefore $\overline{Y} \subset Y$, i.e. $Y = \overline{Y}$.

Conversely, let $Y = \overline{Y}$. Then if $x \overline{\in} Y$, x is not a point of contact of Y, i.e. some neighbourhood U_x of x does not intersect with Y, that is $U_x \subset X \setminus Y$. Thus, the set $X \setminus Y$ can be represented as a union of open sets U_x and is therefore an open set. Theorem 4 is proved.

Theorem 5. *The closure $\overline{Y}$ of an arbitrary set Y of a topological space X is a closed set.*

Proof. According to Theorem 4 we need to prove that $\overline{\overline{Y}} = \overline{Y}$. The inclusion $\overline{Y} \subset \overline{\overline{Y}}$ is obvious, and it remains to verify the inverse inclusion $\overline{\overline{Y}} \subset \overline{Y}$. Let $x \in \overline{\overline{Y}}$. This means that any neighbourhood U of a point x has a non-empty intersection with $\overline{Y}$. For example, if a point $y \in U \cap \overline{Y}$, then the set U is a neighbourhood of y. Since $y \in \overline{Y}$, the set U has a non-empty intersection with the set Y. Hence, x is a point of contact of Y, i.e. $x \in \overline{Y}$. We have demonstrated therefore that $\overline{\overline{Y}} \subset \overline{Y}$. Theorem 5 is proved.

Example 1. Consider a set X consisting of only one element, x. Then on X we can define a unique topology such that its open sets are represented by X and the empty set.

Example 2. Consider a set X consisting of two elements $x \neq y$. In this case several distinct topologies are valid on X. A first topology is represented by the set of all subsets considered as open sets, i.e. $\{\varnothing, \{x\}, \{y\}, X\}$. A second topology is given by the following family of open sets: $\{\varnothing, X\}$. Finally, a third topology can be introduced with $\{\varnothing, \{x\}, X\}$ as open sets. These are all distinct topologies on the same set X, and define three distinct topological spaces.

Example 3. Let X be an arbitrary set. Introduce on X a topology, assuming any subset of X to be open. Then any single-point subset is open and therefore any subset, being the union of its points, is also open. This topology is called *discrete*.

Let X be a topological space and $Y \subset X$ be its subset. Then on Y we can also construct a topology by considering any set $Y \cap U$, where U is open in X, as an open set. In this case the topological space Y is called a *subspace* of the topological space X, and the topology in Y is called an *induced topology*. If X is a metric space and Y is its subspace, a topology in Y can be defined irrespective of the order of operations: restriction of the metric and transition to topology or transition to topology and induction of the metric.

Let $Y \subset X$ be a subset in a topological space X. The set Y is called *dense (everywhere dense)* if $\overline{Y} = X$.

Theorem 6. *If Y_1, $Y_2 \subset X$ are two open dense sets, their intersection $Y = Y_1 \cap Y_2$ is open and dense in X.*

Proof. Let $x \in X$ be an arbitrary point and U its neighbourhood. Since the set Y_1 is dense, we have $U \cap Y_1 \neq \varnothing$, i.e. we can find

a point y such that $y \in U \cap Y_1$. Since $U \cap Y_1$ is an open set and Y_2 is dense, we have $U \cap Y_1 \cap Y_2 \neq \varnothing$, i.e. $U \cap Y \neq \varnothing$, whence it follows that Y is a dense set in X. The theorem is proved.

2.1.3. CONTINUOUS MAPPINGS

The concept of a topological space has been so successfully formulated that the definition of a continuous mapping can be borrowed word for word from mathematical analysis.

Definition according to Cauchy. The mapping $f: X \to Y$ of a topological space is said to be *continuous* at a point $x_0 \in X$ if for any neighbourhood $V(f(x_0))$ of the point $f(x_0) \in Y$ there exists a neighbourhood $U(x_0)$ of $x_0 \in X$ such that $f(U(x_0)) \subset V(f(x_0))$. The mapping f is called *continuous* if it is continuous at each point of space X.

Theorem 7. *The mapping $f: X \to Y$ is continuous if and only if one of the following equivalent conditions is satisfied:*

(a) *the inverse image of any open set is an open set.*

(b) *the inverse image of any closed set is a closed set.*

Proof. Since we have for inverse images $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$, conditions (a) and (b) are equivalent. Suppose f is a continuous mapping and $V \subset Y$ is an open set. We now demonstrate that the inverse image $f^{-1}(V)$ is open. Let $x \in f^{-1}(V)$, then $f(x) \in V$, i.e. V is a neighbourhood of point $f(x)$. By the definition of a continuous mapping, there exists a neighbourhood U of a point x such that $f(U) \subset V$, i.e. $U \subset f^{-1}(V)$, and therefore the set $f^{-1}(V)$ is open. Conversely, let condition (a) be satisfied. If $V \ni f(x_0)$ is a neighbourhood of point $f(x_0)$, then $U = f^{-1}(V) \ni x_0$ is a neighbourhood of x_0 and $f(U) = f(f^{-1}(V)) = V \subset V$. Hence, f is a continuous mapping. The theorem is proved.

Conditions (a) and (b) of Theorem 7 are especially convenient for verifying the continuity of mappings of topological spaces. Below we shall need the following useful theorem.

Theorem 8. *Let a topological space X be defined as a union of its two closed subsets $X = F_1 \cup F_2$ and let $f: X \to Y$ be a mapping of X into a topological space Y. The mapping f is continuous if and only if the restrictions of f to the subsets F_1 and F_2 (denoted as $f \mid F_1: F_1 \to Y$ and $f \mid F_2: F_2 \to Y$) are continuous.*

Proof. Let f be a continuous mapping. We now demonstrate that, say, $f \mid F_1$ is a continuous mapping. The inverse images for $f \mid F_1$ can be calculated by the formula $(f \mid F_1)^{-1}(A) = F_1 \cap f^{-1}(A)$. Thus, if A is a closed set, $f^{-1}(A)$ and $F_1 \cap f^{-1}(A)$ are also closed. Conversely, let $f \mid F_1$ and $f \mid F_2$ be continuous mappings. We shall prove that f is continuous. Suppose $A = \bar{A} \subset Y$. Then,

$$f^{-1}(A) = f^{-1}(A) \cap X = f^{-1}(A) \cap (F_1 \cup F_2)$$
$$= (f^{-1}(A) \cap F_1) \cup (f^{-1}(A) \cap F_2) = (f \mid F_1)^{-1}(A) \cup (f \mid F_2)^{-1}(A).$$

Since the set $(f \mid F_1)^{-1}(A)$ is closed in the subspace F_1 and $F_1 = \overline{F}_1$, we have $(f \mid F_1)^{-1}(A)$ closed in X. Similarly, the set $(f \mid F_2)^{-1}(A)$ is also closed in X, whence it follows that $f^{-1}(A)$ is closed in X. Theorem 8 is proved.

Let us consider a continuous mapping $f\colon X \to Y$ of a topological space X into a topological space Y. If the mapping f is one-to-one and the inverse mapping f^{-1} is also continuous, f is called a *homeomorphism*, and the topological spaces X and Y are called *homeomorphic*. A homeomorphism defines a one-to-one correspondence not only between points of topological spaces X and Y, but also between the topologies themselves, i.e. between families of open and closed sets. A property of a topological space is said to be *topologically invariant* if this property is the same for homeomorphic topological spaces. Thus, while studying topologically invariant properties, we need not distinguish between homeomorphic spaces.

Example 4. The continuity of the mapping $f\colon X \to Y$ is a topologically invariant property. Indeed, if $\varphi\colon X' \to X$ and $\psi\colon Y \to Y'$ are homeomorphisms, the composition $X' \xrightarrow{\psi f \varphi} Y'$ is also a continuous mapping. In general, the composition of two continuous mappings is a continuous mapping.

Example 5. A continuous function, i.e. a continuous mapping of a topological space X into the space $\mathbf{R}^1$ of real numbers, is an important particular case of continuous mappings. The condition that a function f is continuous can be formulated as follows: for any point $x_0 \in X$ and any $\varepsilon > 0$ there exists a neighbourhood U of x_0 such that for $y \in U$ the inequality $|f(x_0) - f(y)| < \varepsilon$ is satisfied. The uniform limit of a sequence of continuous functions can be defined for functions on a topological space. Just like in the case of a function of one real variable, the following statement is valid:

If $f = \lim\limits_{n\to\infty} f_n$ and the sequence $\{f_n\}$ of continuous functions on a topological space X converges uniformly to f, the function f is continuous. Indeed, let $x_0 \in X$ and $\varepsilon > 0$. We can choose a number n such that $|f(x) - f_n(x)| < \varepsilon/3$ for any $x \in X$ (uniform convergence). Further, since the f_n are continuous, there exists a neighbourhood U of x_0 such that $|f_n(x_0) - f_n(x)| < \varepsilon/3$ for $x \in U$. Then, $|f(x_0) - f(x)| < |f(x_0) - f_n(x_0)| + |f_n(x_0) - f_n(x)| + |f_n(x) - (x)| < 3\cdot\varepsilon/3 = \varepsilon$.

Example 6. Let $f\colon X \to Y$ be a continuous mapping of metric spaces and let ρ_1, ρ_2 be metrics on X and Y, respectively. Then the condition that f is continuous can be formulated as follows: *for any $x_0 \in X$ and $\varepsilon > 0$ there exists a $\delta > 0$ such that the inequality $\rho_1(x, x_0) < \delta$ implies $\rho_2(f(x), f(x_0)) < \varepsilon$.*

It would also be useful to extend the concept of the convergence of a numerical sequence to a metric space. We say that the sequence of points $\{x_n\}$ converges to a point x_0, $\lim\limits_{n\to\infty} x_n = x_0$, provided

$\lim_{n\to\infty} \rho(x_0, x_n) = 0$. Many properties of spaces and mappings can be formulated in terms of convergent sequences of a metric space. For example, a set $Y \subset X$ is closed if for any convergent sequence $\{x_n\} \subset Y$ the limit $x_0 = \lim_{n\to\infty} x_n$ also belongs to Y. Further, the condition that the mapping $f\colon X \to Y$ of metric spaces is continuous can be formulated following Heine: *the mapping f is continuous at a point x_0 if the equality $x_0 = \lim_{n\to\infty} x_n$ implies $\lim_{n\to\infty} f(x_n) = f(x_0)$.*

Example 7. Consider two topological spaces X and Y and form a new topological space $X \times Y$. $X \times Y$ is the set of all pairs (x, y), $x \in X$, $y \in Y$, and is called the *Cartesian product of sets* X and Y. Let us now define a topology in $X \times Y$. The set $U \subset X \times Y$ is called open if U can be represented as the union $U = \bigcup_\alpha (V_\alpha \times W_\alpha)$, where $V_\alpha \subset X$, $W_\alpha \subset Y$ are open sets. It is a trivial matter to verify the properties of an open set. The set $X \times Y$ with the topology just defined is called the *Cartesian product of topological spaces* X and Y (these spaces are called *factors* of Cartesian product $X \times Y$). The Cartesian product possesses the following properties: (a) the spaces $X \times Y$ and $Y \times X$ are homeomorphic, and (b) the spaces $(X \times Y) \times Z$ and $X \times (Y \times Z)$ are homeomorphic. In (a) we should choose the homeomorphism in the form $\varphi\colon X \times Y \to Y \times X$, $\varphi(x, y) = (y, x)$. If $U = \bigcup_\alpha (V_\alpha \times W_\alpha)$ is an open set of space $X \times Y$, then $\varphi(U) = \bigcup_\alpha (W_\alpha \times V_\alpha)$ is an open set of $Y \times X$. In (b) we should consider the homeomorphism in the form $\varphi\colon (X \times Y) \times Z \to X \times (Y \times Z)$, $\varphi((x, y,) z) = (x, (y, z))$. The projection of the Cartesian product $X \times Y$ onto one of the factors, say X, $f\colon X \times Y \to X$, $f(x, y) = x$, is a continuous mapping. Indeed, the inverse image of the open set $U \subset X$ is $f^{-1}(U) = U \times Y$, i.e. $f^{-1}(U)$ is an open set.

Example 8. Let X and Y be metric spaces, then the Cartesian product $X \times Y$ admits a metric consistent with the topology of this product. Let ρ_1 and ρ_2 be the metrics of X and Y, respectively. The metric ρ in the Cartesian product $X \times Y$ is defined as

$$\rho((x_1, y_1), (x_2, y_2)) = \max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}. \tag{1}$$

Let us verify properties (a), (b), and (c) of Section 1.1 for the metric. If $\rho((x_1, y_1), (x_2, y_2)) = 0$, then $\rho_1(x_1, x_2) = \rho_2(y_1, y_2) = 0$, i.e. $x_1 = x_2$, $y_1 = y_2$, whence $(x_1, y_1) = (x_2, y_2)$. Conversely, if $(x_1, y_1) = (x_2, y_2)$, then $x_1 = x_2$, $y_1 = y_2$ and hence $\rho_1(x_1, x_2) = \rho_2(y_1, y_2) = 0$, that is, $\rho((x_1, y_1), (x_2, y_2)) = \max\{0, 0\} = 0$. Property (a) is proved. Further,

$$\begin{aligned}\rho((x_1, y_1), (x_2, y_2)) &= \max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}\\ &= \max\{\rho_1(x_2, x_1), \rho_2(y_2, y_1)\}\\ &= \rho((x_2, y_2), (x_1, y_1)),\end{aligned}$$

and property (b) follows. Finally, let us verify the triangle inequality (c):

$$\begin{aligned}\rho((x_1, y_1), (x_2, y_2)) &= \max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}\\ &\leqslant \max\{\rho_1(x_1, x_3) + \rho_1(x_3, x_2), \rho_2(y_1, y_3)\\ &\quad + \rho_2(y_3, y_2)\}\\ &\leqslant \max\{\rho_1(x_1, x_3), \rho_2(y_1, y_3)\}\\ &\quad + \max\{\rho_1(x_3, x_2) + \rho_2(y_3, y_2)\}\\ &= \rho((x_1, y_1), (x_3, y_3))\\ &\quad + \rho((x_3, y_3), (x_2, y_2)).\end{aligned}$$

Thus, formula (1) defines a metric in the Cartesian product $X \times Y$. We now demonstrate that the topology defined by ρ coincides with the topology of Cartesian product. Consider a ball neighbourhood of radius ε with the centre at point (x_0, y_0)

$$\begin{aligned}O_\varepsilon(x_0, y_0) &= \{(x, y): \rho((x, y), (x_0, y_0)) < \varepsilon\}\\ &= \{(x, y): \rho_1(x, x_0) < \varepsilon, \rho_2(y, y_0) < \varepsilon\} = O_\varepsilon(x_0) \times O_\varepsilon(y_0).\end{aligned}$$

Hence, any open set in the sense of metric ρ is open in the topology of a Certasian product and vice versa.

There are some other ways of defining a metric in the Cartesian product. A rather customary method is to establish an analogy between the factors X and Y of the Cartesian product and coordinate axes on a plane. The metric is given by

$$\rho'((x_1, y_1), (x_2, y_2)) = \sqrt{\rho_1(x_1, x_2)^2 + \rho_2(y_1, y_2)^2}. \tag{2}$$

Verification of properties (a), (b), and (c) is left to the reader. In order to prove that the topology defined by metric (2) coincides with that defined by metric (1), it suffices to find constants $C_1 > 0$ and $C_2 > 0$ such that

$$\begin{aligned}C_1\rho((x_1, y_1), (x_2, y_2)) &\leqslant \rho'((x_1, y_1), (x_2, y_2))\\ &\leqslant C_2\rho((x_1, y_1), (x_2, y_2)).\end{aligned} \tag{3}$$

Indeed, if O_ε is a ball neighbourhood in the metric ρ and O'_ε is a ball neighbourhood in the metric ρ', then it is sufficient that the inclusion

$$O_\varepsilon((x_0, y_0)) \subset O'_{\varepsilon'}((x_0, y_0)) \subset O_{\varepsilon''}((x_0, y_0)), \tag{4}$$

be valid at any point, where ε, ε' are chosen for an arbitrary $\varepsilon'' > 0$. From inequality (3) it follows that inclusion (4) is satisfied for $\varepsilon' = C_2\varepsilon''$ and $\varepsilon = \varepsilon'/C_1$, and it remains to prove inequalities (3).

We have

$$\rho'((x_1, y_1), (x_2, y_2)) = \sqrt{\rho_1(x_1, x_2)^2 + \rho_2(y_1, y_2)^2}$$
$$\leqslant \sqrt{2}\max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}$$
$$= \sqrt{2}\rho((x_1, y_1), (x_2, y_2)),$$

$$\rho((x_1, y_1), (x_2, y_2))$$
$$= \max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}$$
$$\leqslant \sqrt{\rho_1(x_1, x_2)^2 + \rho_2(y_1, y_2)^2} = \rho'((x_1, y_1), (x_2, y_2)).$$

Hence, inequalities (3) hold for $C_1 = 1$ and $C_2 = \sqrt{2}$. We have thus demonstrated that in the Cartesian product $X \times Y$ two metrics, (1) and (2), induce the same topology.

Problems

1. Give an example of a metric on a finite set which cannot be induced by any embedding of the set in a Euclidean space.
2. Demonstrate that a finite set on a straight line is closed.
3. Demonstrate that $\rho(x, Y) = \rho(x, \overline{Y})$.
4. Prove that the function $f(x) = \rho(x, Y)$ is continuous for any subset of Y.
5. Demonstrate that any metric on a finite set induces on this set a discrete topology.
6. Prove that an interval, a half-interval, and a segment on the real axis are not pairwise homeomorphic.
7. Prove that the metric $\rho(x, y)$ is a continuous function on the Cartesian square $X \times X$ of a metric space X.
8. Prove that a set X can be represented as the difference of two closed sets if and only if $(\overline{X} \setminus X)$ is a closed set.
9. Demonstrate that under a continuous mapping the image of a dense subset is dense in the image.

2.2. CONNECTEDNESS. SEPARATION AXIOMS

Among all topological spaces we may distinguish narrower classes characterized by certain topological invariants. For instance, in the preceding section (Example 3) we have considered the class of discrete topological spaces. The following property holds true for these spaces: *any mapping f of a discrete space X into a topological space Y is continuous*. In this section we shall consider some other, more important classes of topological spaces.

2.2.1. CONNECTEDNESS

A topological space X is said to be *disconnected* if it contains a subset $Y \subset X$, both open and closed, which is distinct from X and from the empty set $\varnothing$. For example, any discrete topological space consisting of more than one point is disconnected, since any subset of this space is both open and closed. If a space X is disconnected, it can be decomposed into a union of two open, disjoint, non-empty subsets: $X = Y \cup (X \setminus Y)$.

A topological space X is called a *connected topological space* if X cannot be decomposed into a union of two open, disjoint, non-empty subsets.

Example 1. Consider a *segment* $[a, b]$ of the numerical axis and demonstrate that $[a, b]$ is a *connected set*. Suppose the segment $[a, b]$ is represented as a union of two non-empty, open-closed disjoint sets, $[a, b] = A \cup B$. Without loss of generality, we can assume that $a \in A$. Then since A is open, there can be found an $\varepsilon > 0$ such that the half-interval $[a, a + \varepsilon) \subset A$. Let $\varepsilon_0 = \sup \{\varepsilon\}$ with ε running the numbers such that $[a, a + \varepsilon) \subset A$. Then for any $\varepsilon < \varepsilon_0$ the segment $[a, a + \varepsilon) \subset A$, i.e. $a + \varepsilon \in A$ for any $\varepsilon < \varepsilon_0$. Hence, the closeness of A implies that $a + \varepsilon_0 \in A$. The only possibility is $a + \varepsilon_0 = b$, otherwise ε_0 is not equal to $\sup \{\varepsilon\}$. Thus, $[a, b] = A$, i.e. $B = \varnothing$ contrary to the proposition. We can show in a similar fashion that an open interval (a, b) of the numerical axis is a connected set.

We now prove several statements which may be useful in verifying the connectedness of topological spaces.

Theorem 1. *Let* $X = \bigcup_\alpha X_\alpha$, *each* X_α *be connected, and the intersection* $\bigcap_\alpha X_\alpha$ *be non-empty. Then the space* X *is connected.*

Proof. Assume the converse, i.e. $X = A \cup B$, $A \cap B = \varnothing$, and the sets A and B are open and non-empty. Then $X_\alpha = (X_\alpha \cap A) \cup (X_\alpha \cap B)$. Since X_α is connected, then either $X_\alpha \cap A = \varnothing$ or $X_\alpha \cap B = \varnothing$, i.e. each set X is entirely contained either in A or in B. Non-emptiness of A and B implies that there exist points $a \in A$ and $b \in B$. Let $a \in X_{\alpha_0}$ then $X_{\alpha_0} \subset A$; let $b \in X_{\alpha_1}$, then $X_{\alpha_1} \subset B$. Hence, $X_{\alpha_0} \cap X_{\alpha_1} = \varnothing$, which contradicts the condition. Theorem 1 is proved.

Theorem 2. *If in a topological space* X *there exists a connected set* C_{xy} *containing any two points* $x, y \in X$. *Then* X *is a connected topological space.*

Proof. Assume the converse, i.e. $X = A \cup B$, $A \cap B = \varnothing$, and the sets A and B are open and non-empty. Then we can find points $a \in A$ and $b \in B$. Consider the decomposition $C_{ab} = (C_{ab} \cap A) \cup (C_{ab} \cap B)$. The point a belongs to $C_{ab} \cap A$ and the point b to $C_{ab} \cap B$. Hence, the sets $C_{ab} \cap A$ and $C_{ab} \cap B$ are non-empty, open in

C_{ab}, and do not intersect. We have come to contradiction.

Theorem 3. *The image of a connected space is connected under a continuous mapping.*

Proof. Let $f: X \to Y = f(X)$ be a continuous mapping. The condition $Y = f(X)$ means that an inverse image of a non-empty set is also non-empty. Therefore, if Y is a disconnected space, $Y = A \cup B$, $A \cap B = \varnothing$, and A and B are open and non-empty. Hence, $X = f^{-1}(Y) = f^{-1}(A) \cup f^{-1}(B)$, so the sets $f^{-1}(A)$ and $f^{-1}(B)$ are open, non-empty, and do not intersect. Theorem 3 is proved.

Example 2. It follows from Theorem 3 that any continuous real function $y = f(x)$ defined on a segment $[a, b]$ of the real axis takes

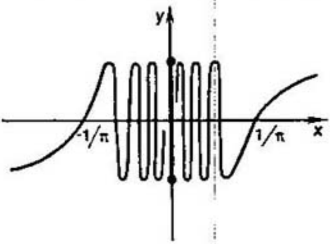

Figure 2.1

intermediate values. If the function f does not assume an intermediate value y_0, the image $f([a, b])$ could be decomposed into the union of two open, non-empty sets: the values which are less than y_0 and those which are larger than y_0, but this contradicts Theorem 3.

A particular case of using Theorems 2 and 3 is the concept of pathwise connectedness. A topological space X is called *pathwise connected* if any two of its points x, y can be joined by a curvilinear segment, i.e. there exists a continuous mapping $f: [0, 1] \to X$ of the segment $[0, 1]$ of the real axis into X such that $f(0) = x$ and $f(1) = y$.

Theorem 4. *A pathwise connected topological space X is connected.*

Proof. Indeed, by Theorem 3 the image $f([0, 1])$ is connected and contains the points x and y. Then by Theorem 2 the space X is connected.

Example 3. Consider a connected topological space which is not pathwise connected. Let X be the closure of the graph of the function $y = f(x) = \sin(1/x)$ as a set in a two-dimensional Euclidean space $\mathbf{R}^2$ (Fig. 2.1). The metric in $\mathbf{R}^2$ is assumed to be classical, in the sense of the length of a segment connecting two points in $\mathbf{R}^2$. Then the set X is a union of the graph of $y = \sin(1/x)$ and the vertical segment $\Gamma_3 = \{(x, y): x = 0, -1 \leqslant y \leqslant +1\}$. The graph of the function f is decomposed into two subsets, each being homeomorphic

to the interval: $\Gamma_1 = \{(x, y): 0 < x < \infty,\ y = f(x)\}$, $\Gamma_2 = \{(x, y): -\infty < x < 0,\ y = f(x)\}$. Thus, if $X = A \cup B$, $A \cap B = \varnothing$, and A and B are open, non-empty sets, then each of the subsets Γ_1, Γ_2, and Γ_3 lies entirely either in A or in B. Let $\Gamma_3 \subset B$. It is a simple matter to verify that any neighbourhood of Γ_3 intersects both Γ_1 and Γ_2, i.e. $\Gamma_1 \subset B$ and $\Gamma_2 \subset B$. Hence, $A = \varnothing$, which contradicts the hypothesis. So X is a connected space.

We now demonstrate that X is not pathwise connected. Consider two points $P = (-1/\pi, 0)$ and $Q = (1/\pi, 0)$ in X. Suppose there exists a continuous mapping f: $[0, 1] \to X$, $f(0) = P$, $f(1) = Q$. The mapping f is defined by two continuous numerical functions: $f(t) = ((x(t), y(t)), y(t) = \sin(1/x(t))$ for $x(t) \neq 0$. Since $x(0) = -1/\pi$, the exact lower bound t_0 stricly exceeds zero for t satisfying $x(t) = 0$. Thus, the conditions $x(t) < 0$, $y(t) = \sin(1/x(t))$ hold on the interval $[0, t_0)$. Since $x(t)$ is a continuous function and there exists a sequence $t_k \geqslant t_0$, $t_k \to t_0$, where $x(t_k) = 0$, we have $x(t_0) = \lim\limits_{t\to t_0} x(t) = 0$. In this case, however, the function $\sin(1/x(t))$ does not have a limit for $t \to t_0 - 0$, and hence the function $y(t)$ is not continuous. Thus, X is not a pathwise connected space.

Example 4. Let us consider two topological spaces X and Y and construct a new topological space $Z = X \cup Y$ consisting of the points of X and Y (we assume here that X and Y do not have common points). By an open set of Z we shall mean any set of the form $W = U \cup V$, where $U \subset X$, $V \subset Y$ are open sets in X and Y. The space Z is called a *topological sum of the spaces* X and Y. The space Z is really disconnected, for its subspaces X and Y are open, disjoint subsets. Using two spaces X and Y, we can construct another topological space called a *connected sum* or a *wedge* of topological spaces. To this end, we fix two points, one in each space: $x_0 \in X$ and $y_0 \in Y$. Consider the union of X and Y and identify in it the points x_0 and y_0. The space thus obtained is denoted $X \vee Y$. By open sets in $X \vee Y$ we shall mean sets U such that the intersections $X \cap U$ and $Y \cap U$ are open in X and Y, respectively. If the spaces X and Y are connected, their connected sum $X \vee Y$ is, by Theorem 1, a connected topological space. If X and Y are pathwise connected spaces, their connected sum $X \vee Y$ is a pathwise connected space.

2.2.2. SEPARATION AXIOMS

A topological space X is called a *Hausdorff space* if for any points $x, y \in X$, $x \neq y$ there exist disjoint neighbourhoods $U(x)$ and $U(y)$, i.e. $U(x) \cap U(y) = \varnothing$. Not every topological space is a Hausdorff one. In the preceding section (Example 2) we have considered the space X consisting of two points x and y the open sets of which are $\{\varnothing, X, \{x\}\}$. The points x, y are then distinct, and no intersecting neighbourhoods of these points exist. In

a Hausdorff space any point $x \in X$ is a closed set. Indeed, if $y \neq x$, we can find a neighbourhood $U(y)$ of point y which does not contain x. Then the set $X \setminus \{x\} = \bigcup_{y \neq x} U(y)$ is open and its complement consisting of one point x is closed.

A discrete topological space X is a Hausdorff space. Indeed, any point $x \in X$ is an open set, and hence for $x \neq y$ the neighbourhoods $\{x\}$ and $\{y\}$ are disjoint.

Theorem 5. *Let X and Y be Hausdorff topological spaces. Then the Cartesian product, the topological sum, and the connected sum of X and Y are Hausdorff topological spaces.*

Proof. Consider two points $(x_1, y_1) \in X \times Y$ and $(x_2, y_2) \in X \times Y$. If these points are distinct, then either $x_1 \neq x_2$ or $y_1 \neq y_2$. In the first case, since X is a Hausdorff space, there exist two disjoint neighbourhoods $U(x_1)$ and $U(x_2)$. Then the neighbourhoods $U(x_1) \times Y$ and $U(x_2) \times Y$ of the points (x_1, y_1) and (x_2, y_2) are disjoint either. In the second case, where Y is a Hausdorff space, there exist disjoint neighbourhoods $V(y_1)$ and $V(y_2)$. Therefore, in this case the neighbourhoods $X \times V(y_1)$ and $X \times V(y_2)$ of the points (x_1, y_1), (x_2, y_2) are disjoint.

Let us consider a pair of distinct points $x, y \in X \cup Y$. If these points lie in the same space, say X, the condition that X is a Hausdorff space implies that there exist disjoint neighbourhoods $U(x)$ and $U(y)$ which are at the same time open sets in the union $X \cup Y$. If the points belong to distinct spaces, say $x \in X$ and $y \in Y$, the spaces X and Y themselves are disjoint neighbourhoods of these points.

Consider a wedge $X \vee Y$ and a pair of distinct points x, y in it. First, let the two points belong to X. Then one of them, say x, does not coincide with a fixed point $x_0 = y_0$, so that disjoint neighbourhoods $U(x)$ and $U(y)$ exist in X. After eliminating x_0 from $U(x)$, we may consider that $U(x)$ does not contain this point. The set $U(x)$ is therefore open in $X \vee Y$. The set $U(y)$ is also open in $X \vee Y$, provided it does not contain the point x_0. If $x_0 \in U(y)$, the set $U(y) \cup Y$ is open in $X \vee Y$ and does not intersect with $U(x)$. This completes the proof of Theorem 5.

Definition. A topological space X is called *normal* if it is a Hausdorff space and for any two disjoint closed sets F_1 and F_2 there exist disjoint neighbourhoods $U_1 \supset F_1$ and $U_2 \supset F_2$.

Normal spaces are the most common occurrence among topological spaces. This class is fairly wide and includes all metric spaces.

Theorem 6. *Any metric space is normal.*

Proof. Let X be a metric space with metric ρ and let F_1 and F_2 be disjoint closed sets. Suppose $x \in F_1$, $\varepsilon(x) = \frac{1}{3}\rho(x, F_2)$ and put $U_1 = \bigcup_{x \in F_1} O_{\varepsilon(x)}(x)$. Similarly, $U_2 = \bigcup_{y \in F_2} O_{\varepsilon'(y)}(y)$, where $\varepsilon'(y) =$

$\frac{1}{3}\rho(y, F_1)$. We have obtained neighbourhoods of the sets F_1 and F_2. We now demonstrate that U_1 and U_2 are disjoint. Assume the converse, i.e. there exists a point $z \in U_1 \cap U_2$. Then for certain points $x \in F_1$ and $y \in F_2$ we have $z \in O_{\varepsilon(x)}(x)$, $z \in O_{\varepsilon'(y)}(y)$, i.e. $\rho(x, z) < \frac{1}{3}\rho(x, F_2)$, $\rho(y, z) < \frac{1}{3}(y, F_1)$. In particular, $\rho(x, z) < \frac{1}{3}\rho(x, y)$, $\rho(y, z) < \frac{1}{3}\rho(y, x)$. Adding the last two inequalities, we obtain $\rho(x, z) + \rho(y, z) < \frac{2}{3}\rho(x, y)$, in contradiction with the triangle inequality. Theorem 6 is proved.

The system of open sets $\{U_\alpha\}$ of a topological space X is called an *open covering* if $X = \bigcup_\alpha U_\alpha$. Covering is a convenient concept for studying topological spaces. For example, if a continuous function f_α is defined in every U_α and the functions f_α and f_β coincide on every intersection $U_\alpha \cap U_\beta$, there exists on X a continuous function f which coincides with f_α in every open set U_α.

Given two open coverings $\{U_\alpha\}$ and $\{V_\beta\}$ of a topological space X. We say that the covering $\{V_\beta\}$ *refines the covering* $\{U_\alpha\}$ or is *finer than* $\{U_\alpha\}$ if each set V_β is contained in the set U_α, $\alpha = \alpha(\beta)$.

Theorem 7. *Let X be a normal topological space and $\{U_\alpha\}_{\alpha=1}^N$ be a finite open covering. Then there exists a finer covering $\{V_\alpha\}_{\alpha=1}^N$ such that $\overline{V}_\alpha \subset U_\alpha$.*

Proof. The theorem implies the existence of a finer covering whose elements are numbered by the same index α and the sets are also included with the index α. Consider a closed set $X \setminus \bigcup_{\alpha=2}^N U_\alpha$ contained in U_1. Since the space X is normal, there exists a neighbourhood V_1 such that $X \setminus \bigcup_{\alpha=2}^N U_\alpha \subset V_1 \subset \overline{V}_1 \subset U_1$. Then the system of sets $\{V_1, U_2, \ldots, U_N\}$ covers X, and we can find a set $V_2 \subset \overline{V}_2$ where the system $\{V_1, V_2, U_3, \ldots, U_N\}$ also covers the space X. Replacing U_k successively by $V_k \subset \overline{V}_k \subset U_k$, we arrive, after N steps, at the covering in question $\{V_1, \ldots, V_N\}$.

Problems

1. Prove that if a finite (countable) number of points is eliminated from $\mathbf{R}^n$ $(n \geqslant 2)$, the remaining space is connected.

2. Prove that if finitely many subspaces of dimension less than $(n-1)$ is eliminated from $\mathbf{R}^n$, the remaining space is connected.

3. Find the maximal number of connectedness components into which $\mathbf{R}^2$ is subdivided by finitely many straight lines. What is the minimal number of the connectedness components?

4. Let $f\colon X \to X$ be a continuous mapping of a Hausdorff space. Prove that the set of fixed points (i.e. $f(x) = x$) is closed.

5. Prove that the image of a normal space mapped continuously into a Hausdorff space is normal.

6. Prove that X is a Hausdorff space if and only if the diagonal $\Delta = \{(x, y)\colon x = y\} \subset X \times X$ is closed in $X \times X$.

7. Prove that the mapping $f\colon X \to Y$ into a Hausdorff space Y is continuous if and only if the graph $\Gamma_f = \{(x, f(x))\colon x \in X\} \subset X \times Y$ is closed in $X \times Y$.

2.3. COMPACT SPACES

Compactness is one of the most important properties of a topological space. In particular, this property is fundamental in the study of real numbers and continuous functions.

2.3.1. DEFINITION

A Hausdorff topological space X is called *compact* if any open covering $\{U_\alpha\}$ has a finite part $\{U_{\alpha_h}\}_{h=1}^{N}$ that covers X. The only condition which can be imposed is that a finite finer covering should exist.

Remark. Such a space was first called *bicompact*, while the term compact space was used for a metric bicompact space. Recently, however, a bicompact space has been called simply a compact space. In what follows we shall use this, more recent, terminology.

Example 1. A real segment $[a, b]$ is a compact space. Indeed, if $\{U_\alpha\}$ is an open covering of $[a, b]$, we can assume, without loss of generality, that each element of U_α is an interval (c_α, d_α) (except for two half-intervals $([a, a')$ and $(b', b])$. Consider the set P of all numbers $x \in [a, b]$ such that the segment $[a, x]$ is covered with finitely many elements of U_α. Then, if $x \in P$, $y < x$, the point y also belongs to P. Since the point a belongs to the element of the covering $[a, a')$, $a' > a$, the set P consists of more than one point a. Let x_0 be the exact upper bound of P, $x_0 > a$. If $x_0 < b$, the point x_0 lies in the interval $U_{\alpha_0} = (c_{\alpha_0}, d_{\alpha_0})$, i.e. $c_{\alpha_0} < x_0 < d_{\alpha_0}$. Suppose y, z are such that $c_{\alpha_0} < y < x_0 < z < d_{\alpha_0}$. Then $y \in P$ and the segment $[a, y]$ is covered with finitely many sets U_β, so that the segment $[a, z]$ is also covered with a finite number of sets U_β. Hence, x_0 is not the upper bound of P. Thus, $x_0 = b$, and $b' < x_0 = b$. If $b' < y < x_0$, then $y \in P$ and again the segment $[a, y]$ is covered with finitely many sets U_β, so that the whole segment $[a, b]$ is also covered with finitely many sets U_β.

2.3.2. THE PROPERTIES OF COMPACT SPACES

Theorem 1. *A compact topological space is normal.*

Proof. Let X be a compact space and F be a closed set in X. Demonstrate that if a point x does not belong to F, there exist disjoint neighbourhoods $U(x)$ and $U(F)$.

Lemma 1. *A closed subset F in a compact space X is a compact space.*

Indeed, if $\{V_\alpha\}$ is an open covering of the set F, each $V_\alpha = U_\alpha \cap F$, where U_α are open sets in X. Then the family of open sets $\{U_\alpha, X \setminus F\}$ covers the space X and hence there exists a finite family $\{U_{\alpha_k}, X \setminus F\}$ covering X. Thus, $\{V_{\alpha_k}\}$ covers F.

Let us proceed with proving Theorem 1. Sinle $x \bar{\in} F$, for any point $y \in F$ there exist disjoint neighbourhoods $U_y \ni x$ and $V_y \ni y$ because X is a Hausdorff space. Obviously, the family $\{V_y\}$ covers the set F. According to Lemma 1, we can find finitely many sets $\{V_{y_k}\}_{k=1}^{N}$ covering F, so that the neighbourhood $\bigcup\limits_{k=1}^{N} U_{y_k}$ of point x does not intersect with the union $\bigcup\limits_{k=1}^{N} V_{y_k} \supset F$.

Let us consider two disjoint closed sets F_1 and F_2 in a compact space X. Then for any point $x \in F_1$ we can find disjoint neighbourhoobs U_x and V_x of the point x and of the set F_2, respectively. The family $\{U_x\}$ covers the compact set F_1; hence, there exists a finite covering $\{U_{x_k}\}_{k=1}^{N}$. Thus, the union $\bigcup\limits_{k=1}^{N} U_{x_k}$ contains F_1 and does not intersect with the intersection $\bigcap\limits_{k=1}^{N} V_{x_k}$ containing F_2. Theorem 1 is proved.

Theorem 2. *Let $F \subset X$ be a compact subspace in a Hausdorff topological space X. Then F is a closed set in X.*

Proof. Let x be an arbitrary point of X not belonging to F. Since X is a Hausdorff space, there exist for points $y \in F$ disjoint neighbourhoods $U_y \ni y$ and $V_y \ni x$. Then the family $\{U_y\}$ covers the set F and, as this set is compact, we can find a finite family $\{U_{y_k}\}_{k=1}^{N}$ that covers F. Thus, the union $\bigcup\limits_{k=1}^{N} U_{y_k}$ contains F and does not intersect with $\bigcap\limits_{k=1}^{N} V_{y_k}$. Hence, the point x has a neighbourhood not intersecting with F, which means that the set F is closed. The theorem is proved.

Theorem 3. *Let $f\colon X \to Y$ be a continuous mapping of a compact space X into a space Y. Then the image $f(X)$ is a compact space.*

Proof. Let $\{U_\alpha\}$ be an open covering of the set $f(X)$. Then the family $\{f^{-1}(U_\alpha)\}$ is an open covering of the compact space X.

Hence, some finite family $\{f^{-1}(U_{\alpha_k})\}_{k=1}^N$ covers X and therefore the family $\{U_{\alpha_k}\}_{k=1}^N$ covers the image $f(X)$. The theorem is proved.

Theorems 2 and 3 imply the following general property of a continuous function on a compact space known from mathematical analysis.

Theorem 4. *Let $f\colon X \to \mathbf{R}^1$ be a continuous function on a compact space X. Then the function f is bounded and attains a maximum (minimum) value.*

Proof. According to Theorem 3, the image $f(X)$ is a compact subspace in $\mathbf{R}^1$ and, according to Theorem 2, $f(X)$ is a closed set. If the image $f(X)$ were unbounded, the system of intervals $U_n = (-n, n)$ would cover $f(X)$ and no finite subcovering could be found in this system. Suppose $A = \sup\limits_{x \in X} \{f(x)\}$. Then there exists a sequence $x_n \in X$ such that $\lim\limits_{n\to\infty} f(x_n) = A$, i.e. $A \in f(X)$, $A = f(x_0)$, $x_0 \in X$. Similarly, $B = \inf\limits_{x \in X} \{f(x)\}$ is also a value of the function f. The theorem is proved.

2.3.3. METRIC COMPACT SPACES

In the case of metric spaces the compactness property can be formulated in terms that are customary for mathematical analysis.

Theorem 5. *Let X be a metric space. The space X is compact if and only if one of the following equivalent properties is satisfied:*

(a) *any sequence $\{x_n\}$ has a convergent subsequence,*

(b) *any sequence of embedded, non-empty, closed subsets $\{F_n\}$, $F_n \supset F_{n+1}$, has a non-empty intersection.*

Proof. Let X be a compact metric space and $\{F_n\}$ a sequence of embedded, non-empty, closed sets, $F_n \supset F_{n+1}$. If $\bigcap F_n = \varnothing$, the open sets $G_n = X \setminus F_n$ cover the space X, and $G_n \subset G_{n+1}$. Since X is a compact space, it is covered with a finite system $\{G_{n_1}, \ldots, G_{n_s}\}$. Without loss of generality we assume that $n_1 \leqslant n_2 \leqslant \ldots \leqslant n_s$. Then $G_{n_k} \subset G_{n_s}$, i.e. $X = G_{n_s}$ and hence $F_{n_s} = \varnothing$, which contradicts the hypothesis.

Let us now demonstrate that property (a) follows from property (b). Suppose $\{x_n\}$ is an arbitrary sequence of points of X. If $\{x_n\}$ does not have limit points, any subset $F \subset \{x_n\}$ is closed. We can therefore construct a sequence of embedded closed subsets $\{F_n\}$ with an empty intersection by putting $F_n = \{x_{n+1}, x_{n+2}, \ldots\}$. Hence, by property (b) the sequence $\{x_n\}$ has a limit point $x_0 \in X$. Choose a subsequence $\{x_{n_k}\}$ convergent to x_0 and consider a neighbourhood $O_1(x_0)$. There exists a number n_1 such that $x_{n_1} \in O_1(x_0)$. Let the terms $\{x_{n_1}, \ldots, x_{n_s}\}$, $n_1 < n_2 < \ldots < n_s$, be already

constructed, and $\rho(x_0, x_{n_k}) < 1/k$. Let also

$$\varepsilon = \min\left\{\frac{1}{s+1}, \rho(x_0, x_1), \ldots, \rho(x_0, x_{n_s})\right\}.$$

Then in the neighbourhood $O_\varepsilon(x_0)$ we can find a point $x_{n_{s+1}} \in O_\varepsilon(x_0)$. Clearly, $\rho(x_0, x_{n_{s+1}}) < \varepsilon \leqslant \frac{1}{s+1}$ and hence $n_{s+1} > n_s$. Thus, we have constructed, by induction, the subsequence $\{x_{n_s}\}$ with $\lim\limits_{s\to\infty} \rho(x_0, x_{n_s}) = 0$, which means that $\lim\limits_{s\to\infty} x_{n_s} = x_0$.

Finally, we shall prove that property (a) implies the compactness of X. Let $\{U_\alpha\}$ be an arbitrary open covering of the metric space X. Suppose there exists a number ε such that the covering with balls $\{O_\varepsilon(x): x \in X\}$ is inscribed in $\{U_\alpha\}$. Then ε is called the *Lebesgue number* of the covering $\{U_\alpha\}$.

Lemma 1. *If condition* (a) *of Theorem* 5 *is satisfied, any open covering has a Lebesgue number.*

Proof. For any point $x \in X$ we select a number $\varepsilon(x)$ equal to the exact upper bound of the numbers $\delta > 0$ such that the ball $O_\delta(x)$ lies in an element of the covering U_α. If the function $\varepsilon(x)$ has a strictly positive $\varepsilon_0 = \inf\limits_{x\in X} \varepsilon(x)$, then $\varepsilon_0/2$ is the Lebesgue number of $\{U_\alpha\}$. If $\inf\limits_{x\in X} \varepsilon(x) = 0$, there exists a sequence $\{x_n\}$ such that $\lim\limits_{n\to\infty} \varepsilon(x_n) = 0$. According to property (a), we can choose a subsequence x_{n_k} such that $\lim\limits_{k\to\infty} x_{n_k} = x_0$. Suppose k_0 is a number such that $\rho(x_0, x_{n_k}) < \frac{1}{4}\varepsilon(x_0)$ for $k > k_0$. Then, $O_{\frac{1}{4}\varepsilon(x_0)}(x_{n_k}) \subset O_{\frac{1}{2}\varepsilon(x_0)}(x_0) \subset U_\alpha$. This means that $\varepsilon(x_{n_k}) > \frac{1}{4}\varepsilon(x_0)$ and hence $\lim\limits_{k\to\infty} \varepsilon(x_{n_k}) \geqslant \frac{1}{4}\varepsilon(x_0) > 0$. The lemma is proved.

Let us consider an arbitrary covering $\{U_\alpha\}$. By Lemma 1, the covering with balls $\{O_\varepsilon(x): x \in X\}$ is inscribed in $\{U_\alpha\}$. Suppose $\{U_\alpha\}$ does not have a finite subcovering. Then $\{O_\varepsilon(x): x \in X\}$ does not have a finite subcovering either. Fix an arbitrary point $x_1 \in X$. Since the ball $O_\varepsilon(x_1)$ does not cover the space X, there exists a point $x_2 \overline{\in} O_\varepsilon(x_1)$. Suppose the points $x_1, \ldots, x_n$ are so constructed that $x_k \overline{\in} O_\varepsilon(x_s)$ for $n \geqslant k > s$. The system of balls $\{O_\varepsilon(x_1), \ldots, O_\varepsilon(x_n)\}$ does not cover X, so that there exists a point $x_{n+1} \overline{\in} O_\varepsilon(x_1) \cup \ldots \cup O_\varepsilon(x_n)$. Thus, we have constructed, by induction, the sequence $\{x_n\}$ with $\rho(x_k, x_n) \geqslant \varepsilon > 0$. On the other hand, property (a) implies that there exists a subsequence $\{x_{n_k}\}$ such that $\lim\limits_{k\to\infty} x_{n_k} = x_0$, i.e. $\lim\limits_{k\to\infty} \rho(x_{n_k}, x_0) = 0$. In particular, an appropriate choice of the

number k_0 among $k > k_0$ gives $\rho(x_{n_k}, x_0) < \frac{1}{2}\varepsilon$, whence

$$\rho(x_{n_k}, x_{n_{k+1}}) \leqslant \rho(x_{n_k}, x_0) + \rho(x_{n_{k+1}}, x_0) < \varepsilon.$$

This contradiction proves Theorem 5.

As a corollary of Theorem 5, we formulate the following fundamental property of real numbers: *the sequence of embedded segments of the real axis has a common point.*

2.3.4. OPERATIONS OVER COMPACT SPACES

If a Hausdorff space X is a union of a finite number of its compact subsets, X is a compact space. Indeed, if $X = \bigcup_{k=1}^{n} X_k$, where X_k are compact spaces, and $\{U_\alpha\}$ is an open covering of X, then each subset X_α is covered with finitely many elements $\{U_{\alpha_{k,1}}, \ldots, U_{\alpha_{k,n_k}}\}$. Then the entire finite system $\{U_{\alpha_{k,s}}, 1 \leqslant k \leqslant n, 1 \leqslant s \leqslant n_k\}$ covers the whole space X. The following theorem is very important.

Theorem 6. *The Cartesian product $X \times Y$ of compact spaces X and Y is a compact space.*

Proof. By Theorem 5 of Sec. 2.2, the Cartesian product $X \times Y$ is a Hausdorff space. Let $\{U_\alpha\}$ be an arbitrary covering of $X \times Y$. Without loss of generality, we may assume that $U_\alpha = V_\alpha \times W_\alpha$. Fix a point $x \in X$ and consider a subspace $x \times Y \subset X \times Y$. The subspace $x \times Y$ is homeomorphic to Y and is therefore a compact space. Thus, there exists finitely many elements $\{U_{\alpha(k,x)}\}_{k=1}^{n(x)}$ covering the space $x \times Y$. Put $V'_x = \bigcap_{k=1}^{n(x)} V_{\alpha(k,x)}$. Then the open sets $\{V'_x \times W_{\alpha(k,x)}\}$ are refinements of the covering $\{U_\alpha\}$ and cover $x \times Y$. The sets $\{W_{\alpha(k,x)}\}_{k=1}^{n(x)}$ cover the space Y, $\bigcup_{k=1}^{n(x)} W_{\alpha(k,x)} = Y$, and the sets $\{V'_x\}$ cover X. Since the space X is compact, it is covered with a finite family $\{V'_{x_s}\}_{s=1}^{m}$. Then the space $X \times Y$ is covered with a finite family $\{V'_{x_s} \times W_{\alpha(k,x_s)} : 1 \leqslant s \leqslant m, 1 \leqslant k \leqslant n(x_s)\}$. The theorem is proved.

Remark. If the spaces X and Y are metric, the proof of Theorem 6 is much simpler. Indeed, consider a sequence $\{z_n\}$, $z_n = (x_n, y_n)$. Since X is a compact space, there exists, by Theorem 5, a convergent subsequence $\{x_{n_k}\}$. The compactness of Y implies that in the sequence $\{y_{n_k}\}$ we extract a convergent subsequence $\{y_{n_{k_s}}\}$. Thus, the subsequence $\{z_{n_{k_s}}\}$ is also convergent.

Problems

1. Prove that a union of finitely many compact spaces is compact.
2. Let X be a metric, non-compact space. Prove that there exists a continuous function unbounded on this space.
3. Prove that a metric compact space has a countable dense subset.
4. Given two closed disjoint sets A and B in a compact metric space. Prove that $\rho(A, B) > 0$.

2.4. FUNCTIONAL SEPARABILITY. PARTITION OF UNITY

In this section we shall prove the theorems which enable continuous functions to be analyzed under relaxed conditions, as in the case of functions of one real variable. We have already noted that a continuous function on a topological space behaves, in many respects, like a function of one real variable: namely, the sum $f + g$, product $f \cdot g$, and ratio f/g $(g \neq 0)$ of two continuous functions are also continuous functions. A limiting process is valid in the class of continuous functions on a topological space X. A sequence of functions f_n *converges uniformly* to a function f if for any $\varepsilon > 0$ there exists a number N such that for $n > N$ the inequality $|f(x) - f_n(x)| < \varepsilon$ is satisfied for any point $x \in X$.

Theorem 1. *The uniform limit of a sequence of continuous functions on a topological space X is a continuous function.*

Proof. Let $f(x) = \lim\limits_{n\to\infty} f_n(x)$. We now demonstrate that the function f is continuous. Fix a number $\varepsilon > 0$ and a point $x_0 \in X$. Then there exists a number n such that for any point $x \in X$ we have $|f(x) - f_n(x)| < \varepsilon/3$. Since f_n is a continuous function, there exists a neighbourhood $O(x_0)$ such that for $x \in O(x_0)$ the inequality $|f_n(x) - f_n(x_0)| < \varepsilon/3$ is satisfied. Then we have for $x \in O(x_0)$

$$|f(x) - f(x_0)| < |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.$$

Theorem 1 is proved.

2.4.1. FUNCTIONAL SEPARABILITY

Let F_1 and F_2 be two disjoint closed sets in a topological space X. To construct disjoint neighbourhoods of these sets, it suffices to construct a continuous function f on X such that $f(x) \geqslant a$ for $x \in F_1$, $f(x) \leqslant b$ for $x \in F_2$, and $a > b$. Thus, as the neighbourhoods of the sets F_1 and F_2 we can select the inverse images of the intervals, $f^{-1}(c, \infty)$ and $f^{-1}(-\infty, c)$, where $b < c < a$. The state-

ment, converse in a certain sense, is also valid (it is called Urysohn's lemma).

Theorem 2. *Let X be a normal topological space and F_0, F_1 be two closed disjoint sets. Then there exists a continuous function $f\colon X \to [0, 1]$ such that $f \mid F_0 \equiv 0$ and $f \mid F_1 \equiv 1$.*

Proof. Construct a system of open sets Γ_r numbered by all binary-rational numbers $0 \leqslant r \leqslant 1$, such that the following conditions are satisfied: (1) $F_0 \subset \Gamma_0$, (2) $\bar{\Gamma}_1 \subset X \setminus F_1$, (3) $\bar{\Gamma}_r \subset \Gamma_{r'}$ for $r < r'$.

Then the system of open sets $\{\Gamma_r\}$ can be extended by adding the sets $\Gamma_t = \bigcup_{r<t} \Gamma_r$, where t is an arbitrary real number, $0 \leqslant t \leqslant 1$. The system of open sets $\{\Gamma_t\}$ satisfies the same conditions: $\bar{\Gamma}_t \subset \Gamma_{t'}$ for $t < t'$. Indeed, choosing rational numbers r and r' such that $t < r < r' < t'$, we obtain $\Gamma_t \subset \Gamma_r$, $\Gamma_{r'} \subset \Gamma_{t'}$. Hence, $\bar{\Gamma}_t \subset \bar{\Gamma}_r \subset \Gamma_{r'} \subset \Gamma_t$. Construct a function $f\colon X \to [0, 1]$, putting $f(x) = \sup\{t\colon x \bar{\in} \Gamma_t\}$. We now demonstrate that f it continuous. Fix a point x_0 and an $\varepsilon > 0$, and take $t_0 = f(x_0)$. By the definition of f we have $x_0 \bar{\in} \Gamma_{(t_0-\varepsilon/2)}$, $x_0 \in \Gamma_{(t_0+\varepsilon/2)}$. Consider a neighbourhood of x_0 equal to $U = \Gamma_{(t_0+\varepsilon/2)} \setminus \bar{\Gamma}_{(t_0-\varepsilon/2)}$. If $y \in U$, then $y \in \Gamma_{(t_0+\varepsilon/2)}$, $y \bar{\in} \bar{\Gamma}_{(t_0+\varepsilon/2)}$. Owing to the definition of the function f, the inequalities $t_0 - \varepsilon/2 \leqslant f(y) \leqslant t_0 + \varepsilon/2$ are satisfied, i.e. $|f(x_0) - f(y)| < \varepsilon$. Thus, the function f is continuous.

To complete the proof of the theorem, it remains to construct a system of open sets Γ_r satisfying conditions (1), (2), and (3). Since the space X is normal, for any closed set F and its neighbourhood U, $F \subset U$, there exists another neighbourhood V such that $F \subset V \subset \bar{V} \subset U$. For the sake of brevity we shall write $V \Subset U$, provided $V \subset \bar{V} \subset U$. Thus, $F_0 = \bar{F}_0 \subset X \setminus F_1$ and therefore $F_0 \Subset X \setminus F_1$. Hence, there exists an open set Γ_0 such that $F_0 \Subset \Gamma_0 \Subset X \setminus F_1$. Similarly, we can find an open set Γ_1 such that $F_0 \Subset \Gamma_0 \Subset \Gamma_1 \Subset X \setminus F_1$. Suppose that for all binary-rational numbers $r = p/2^n$, $0 \leqslant p \leqslant 2^n$, the open sets Γ_r have already been constructed, and $\Gamma_{(p/2^n)} \Subset \Gamma_{((p+1)/2^n)}$. Define the set $\Gamma_{(2p+1)/2^{n+1}}$ in such a way that $\Gamma_{p/2^n} \Subset \Gamma_{(2p+1)/2^{n+1}} \Subset \Gamma_{(p+1)/2^n}$. Hence, we have constructed, by induction, the whole family $\{\Gamma_r\}$. If $x \in F_0$, then $f(x) = 0$, and if $x \in F_1$, then $f(x) = 1$. Theorem 2 is proved.

The following important theorem is a consequence of Theorem 2.

Theorem 3. *Let X be a normal topological space, $F \subset X$ a closed set, and $f\colon F \to \mathbf{R}^1$ a continuous function on F. Then the function f is extendable to the continuous function $g\colon X \to \mathbf{R}^1$ on the entire space X. If the function f is bounded, $|f(x)| \leqslant A$, the function g can also be chosen in such a manner that it is bounded by the same constant, $|g(x)| \leqslant A$.*

Proof. First, let us assume that f is bounded, $|f(x)| \leqslant A$. Put $\varphi_0(x) = f(x)$ and consider two closed subsets $A_0 = \{x: \varphi_0(x) \leqslant -A/3\}$ and $B_0 = \{x: \varphi_0(x) \geqslant A/3\}$. Since the sets A_0 and B_0 are disjoint, there exists, according to Theorem 2, a continuous function $f_0: X \to [-A/3, A/3]$ which is $-A/3$ on A_0 and $A/3$ on B_0. In other words, $|f_0(x)| \leqslant A/3$, $x \in X$, $|\varphi_0(x) - f_0(x)| \leqslant 2A/3$. Put $\varphi_1(x) = \varphi_0(x) - f_0(x)$. Then the function φ_1 is bounded on F by the constant $2A/3$. Thus, repeating the procedure, we can construct two disjoint closed sets $A_1 = \{x: \varphi_1(x) \leqslant -2A/9\}$ and $B_1 = \{x: \varphi_1(x) \geqslant 2A/9\}$, and a continuous function $f_1: X \to [-2A/9, 2A/9]$ which is $-2A/9$ on A_1 and $2A/9$ on B_1. In other words, $|f_1(x)| \leqslant \frac{2}{3}\cdot\frac{1}{3}A$, $|\varphi_1(x) - f_1(x)| \leqslant 4A/9$. Repeating this process infinitely many times, we can construct two sequences of functions $f_n: X \to \mathbf{R}^1$ and $\varphi_n: F \to \mathbf{R}^1$ satisfying the following conditions: $\varphi_{n+1}(x) = \varphi_n(x) - f_n(x)$,

$$|f_n(x)| \leqslant \left(\frac{2}{3}\right)^n \frac{A}{3}, \tag{1}$$

$$|\varphi_n(x)| \leqslant \left(\frac{2}{3}\right)^n A. \tag{2}$$

Next,

$$\begin{aligned} f(x) &= \varphi_0(x) = f_0(x) + \varphi_1(x) \\ &= \ldots = f_0(x) + f_1(x) + \ldots + f_n(x) + \varphi_{n+1}(x) \ \text{ for } x \in F. \end{aligned}$$

Inequality (1) implies that the series $\sum\limits_{k=0}^{\infty} f_k(x)$ converges uniformly on the entire space X. According to Theorem 1, the function $g(x) = \sum\limits_{k=0}^{\infty} f_k(x)$ is continuous and $g(x) = f(x)$ for $x \in F$. From inequality (1) we have

$$|g(x)| \leqslant \sum_{k=0}^{\infty} |f_n(x)| \leqslant \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^n \frac{A}{3} = A.$$

Theorem 3 is proved for a bounded function f.

Turning to a general case, let us consider a homeomorphism $h: \mathbf{R}^1 \to (-1, 1)$. Then the composition $hf: F \to (-1, 1)$ is a continuous bounded function. Applying Theorem 3 to the bounded function hf, we can construct a continuous function $g: X \to [-1, 1]$ which is an extension of the function hf. The function g assumes the values ± 1 on a closed set F_1 disjoint with the set F. According to Theorem 2, there exists a continuous function $\psi: X \to [0, 1]$ equal to 1 on F and 0 on F_1. Then the function $g_1(x) = \psi(x) g(x)$, $x \in X$, coincides with hf on F and does not assume the values ± 1. Thus

the function g_1 maps X into the interval $(-1, 1)$. Finally, put $g_2(x) = h^{-1}g_1(x)$. Then $g_2(x) = h^{-1}g_1(x) = h^{-1}hf(x) = f(x)$ for $x \in F$. This completes the proof of the theorem.

2.4.2. PARTITION OF UNITY

The *support* of a continuous function on a topological space X is the closure of the set of those points $x \in X$ for which $f(x) \neq 0$. The support of a function f is denoted by supp f. Thus, a function f is identically zero outside the support. A convenient method in topology is the decomposition of a function into a sum in such a way that any term has a sufficiently small support.

Theorem 4. *Let X be a normal space and $\{U_\alpha\}$ a finite open covering. Then there exist functions φ_α: $X \to \mathbf{R}^1$, $0 \leqslant \varphi_\alpha(x) \leqslant 1$, such that:* (a) supp $\varphi_\alpha \subset U_\alpha$, (b) $\sum_\alpha \varphi_\alpha(x) \equiv 1$.

The system of functions $\{\varphi_\alpha\}$ is called a *partition of unity subordinate to the covering* $\{U_\alpha\}$. Here we do not assume that the covering is finite, but only require that any point $x \in X$ should have a neighbourhood $O(x)$ intersecting finitely many supports supp φ_α. In general, the system of subsets $\{A_\alpha\}$ of a topological space X is called a *locally finite system* if for any point $x \in X$ there exists a neighbourhood $O(x)$ which has a non-empty intersection only with finitely many sets A_α.

Example. The covering of the real axis $\mathbf{R}^1$ with intervals $(-n, n)$ is not locally finite. On the contrary, the covering with intervals $(n, n + 2)$ is locally finite.

Thus, if the family of supports $\{\text{supp } \varphi_\alpha\}$ of continuous functions φ_α is locally finite, only finitely many functions φ_α are non-zero at any point $x \in X$ and therefore the sum $\varphi(x) = \sum_\alpha \varphi_\alpha(x)$ is correctly defined at any point $x \in X$. The function thus obtained is continuous. Indeed, the continuity of φ can be verified independently in some neighbourhood of each point. Since the system $\{\text{supp } \varphi_\alpha\}$ is locally finite, there exists a neighbourhood $O(x)$ in which only finitely many functions $\varphi_{\alpha_1}, \ldots, \varphi_{\alpha_N}$ are non-zero. Hence, in the neighbourhood $O(x)$ the infinite sum appearing in the expression for φ is, in fact, finite, i.e. $\varphi(y) = \sum_{k=1}^{N} \varphi_{\alpha_k}(y)$ for $y \in O(x)$. The function φ is therefore continuous at every point $x \in X$.

Proof of Theorem 4. Let us consider a finite covering $\{U_\alpha\}$. According to Theorem 7 of Sec. 2.2, there exists a finer covering $\{V_\alpha\}$ such that $\overline{V}_\alpha \subset U_\alpha$. By Theorem 2, there exists a continuous function ψ_α on X which satisfies the conditions: $\psi_\alpha \mid \overline{V}_\alpha \equiv 1$, $\psi_\alpha \mid (X \setminus U_\alpha) \equiv 0$, $0 \leqslant \psi_\alpha(x) \leqslant 1$. This means that supp $\psi_\alpha \subset U_\alpha$ and

$\psi_\alpha(x) > 0$ for $x \in V_\alpha$. Put $\psi(x) = \sum_\alpha \psi_\alpha(x)$. The function is continuous. We now demonstrate that $\psi(x) > 0$ at each point $x \in X$. Indeed, since the system $\{V_\alpha\}$ covers the space X, a number α_0 can be found such that $x \in V_{\alpha_0}$, i.e. $\psi_{\alpha_0}(x) > 0$. Thus, $\psi(x) = \sum_\alpha \psi_\alpha(x) \geqslant \psi_{\alpha_0}(x) > 0$. Let us finally put $\varphi_\alpha(x) = \psi_\alpha(x)/\psi(x)$. Then,

$$\operatorname{supp} \varphi_\alpha = \operatorname{supp} \psi_\alpha \subset U_\alpha, \quad 0 \leqslant \varphi_\alpha(x) \leqslant 1,$$

and

$$\sum_\alpha \varphi_\alpha(x) = \sum \psi_\alpha(x)/\psi(x) = (\sum \psi_\alpha(x))/\psi(x) = \psi(x)/\psi(x) \equiv 1.$$

Theorem 4 is proved.

Remark. Theorem 4 can be extended to locally finite open covering. To this end, it suffices to prove Theorem 7 of Sec. 2.2 for locally finite coverings. If a space X is compact, we can always confine ourselves to finite coverings, while for non-compact spaces there exist essentially infinite coverings for which we should, nevertheless, construct a partition of unity. Thus, given a locally finite covering that refines a given covering $\{U_\alpha\}$, a partition of unity can be constructed. The following *statement* is formulated without proof.

Let $X \subset \mathbf{R}^n$ be an arbitrary subspace of a Euclidean space and let $\{U_\alpha\}$ be an open (not necessarily finite) covering of X. Then there exists a partition of unity subordinate to $\{U_\alpha\}$.

Problems

1. Prove that in Theorem 2 we may require the smoothness of the function f, provided X is a Euclidean space $\mathbf{R}^n$.
2. Prove that in Theorem 4 we may require the smoothness of the functions φ_α if $X = \mathbf{R}^n$.

Chapter 3

Smooth Manifolds (General Theory)

INTRODUCTION

In Chapter 1 we have shown that a coordinate system describing the position of a point in space is an indispensable tool for studying geometrical objects. Using coordinate systems, we can apply the methods of differential and integral calculus to solve various problems. Therefore, an analysis of spaces which admit such concepts as differentiable or smooth functions and differentiation and integration has emerged as an independent branch of geometry.

Some examples of these spaces may prove useful.

Example 1. Consider a unit circle S^1 on a plane (i.e. in a two-dimensional Euclidean space $\mathbf{R}^2$). To describe points of the circle by coordinates, we define it as a set of points satisfying the equation

$$x^2 + y^2 = 1 \tag{1}$$

where (x, y) are Cartesian coordinates on the plane. Thus, each point $P \in S^1$ is uniquely defined by a pair of numbers, Cartesian coordinates x and y. For points of a circle, however, there is no need of giving both coordinates. If the coordinate x of point P is known, the second coordinate is easily found from equation (1): $y = \pm\sqrt{1 - x^2}$, i.e. the coordinate y is uniquely (to within a sign) related to the coordinate x. Furthermore, if a point $P_0 = (x_0, y_0)$ does not lie on the x-axis, i.e. $y_0 \neq 0$, there exists a sufficiently small neighbourhood U of point P_0 such that for all points $P \in U$ the sign of the coordinate y of P is uniquely determined by the sign of the coordinate y_0 of point P_0. We may therefore say that points of the circle S^1 are described by one numerical parameter, the Cartesian coordinate x. To be more exact, only the points of the upper semicircle, i.e. the points satisfying $y > 0$, are uniquely defined by one numerical parameter x. Similarly, the points of the lower semicircle, i.e. the points satisfying $y < 0$, are also defined by one numerical parameter x. The parameter x varies within the same range, namely on the interval $(-1, 1)$ of the real axis, both for the upper and lower semicircles. If we need to describe parametrically points of a circle, including "singular" points $P_0 = (1, 0)$ and $P_1 =$

$(-1, 0)$, we will have to interchange the coordinates x and y and express x in terms of y through equation (1).

A question naturally arises: can we introduce such a parametrization of points of a circle S^1 that all the points are uniquely defined by a numerical parameter? The parameter φ, equal to the angle between the x-axis and radius vector ending at point P, seems to be the most appropriate approximation, though this parameter is not defined uniquely. And if we consider the values of φ only in some interval, say $0 < \varphi \leqslant 2\pi$, the function associating the angular parameter φ with a point $P \in S^1$ will suffer a discontinuity at point $P_0 = (1, 0)$.

Thus, we arrive at the following *assertion*. No continuous function exists on a circle S^1 (as a topological space) whose values uniquely describe points of the circle.

Example 2. Let us consider in a three-dimensional Euclidean space a two-dimensional sphere S^2 defined by

$$x^2 + y^2 + z^2 = 1. \tag{2}$$

Just like in the case of a circle, it is unnecessary to give all three Cartesian coordinates (x, y, z) for points of the sphere S^2, since one coordinate, say z, can be expressed in terms of the other two: $z = \pm\sqrt{1 - x^2 - y^2}$. Obviously, on the upper (and similarly the lower) hemisphere the coordinate z of a point P is uniquely related to the coordinates x and y.

As before, it can be shown that there exist no two continuous functions on the two-dimensional sphere S^2 such that their values uniquely define a point P on the sphere.

These examples demonstrate that the only possibility is to reject the construction of a unique coordinate system for all points of a space under consideration and use different coordinate systems for different parts of a space. A rigorous analysis of the situation has led in geometry to a specific concept of a manifold.

3.1. THE CONCEPT OF A MANIFOLD

3.1.1. FUNDAMENTAL DEFINITIONS

A metric space M is called an *n-dimensional manifold* (or simply *manifold*) if any point P of the space is contained in a neighbourhood $U \subset M$ homeomorphic to a domain V of a Euclidean space $\mathbf{R}^n$. This condition can be formulated in brief as follows: an n-dimensional manifold M is locally homeomorphic to a domain in a Euclidean space $\mathbf{R}^n$, in which case the dimension of M is said to be equal to n or $\dim M = n$. Thus, if M is an n-dimensional manifold, we can find in M a system of open sets $\{U_i\}$ numbered by finitely (or infinitely) many indices i and a system of homeomorphisms

φ_i: $U_i \to V_i \subset \mathbf{R}^n$ of the sets U_i on the domain V_i. The system $\{U_i\}$ must cover the space M, i.e. $M = \bigcup\limits_i U_i$, and the domains V_i may, in general, intersect one another.

Suppose a Cartesian coordinate system $(x^1, \ldots, x^n)$ is valid in a Euclidean space $\mathbf{R}^n$. Then for any point $P \in U_i$ the Cartesian coordinates of the point $\varphi_i(P) \in V_i$ can be considered as a numerical parametrization of P. The homeomorphism φ_i is therefore called

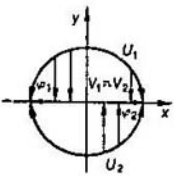

Figure 3.1

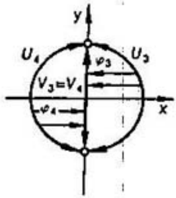

Figure 3.2

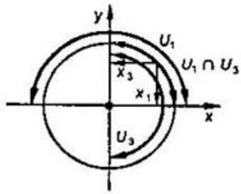

Figure 3.3

a *coordinate homeomorphism* and the Cartesian coordinates $(x^1, \ldots, x^n)$ of $\varphi_i(P)$ are termed *local coordinates of the point* $P \in U_i$ and denoted by $x^k = x^k(P)$*, $k = 1, \ldots, n$. A system of functions $x^k = x^k(P)$* given on an open set U_i is called a *local coordinate system*, and the open set U_i together with a local coordinate system defined on it is called a *chart* of a manifold M. Thus, a chart is a pair (U_i, φ_i), and we shall denote it, for brevity, only by the first symbol, U_i. A set of charts $\{U_i\}$ covering the entire manifold M is called an *atlas*. It is convenient to number local coordinates of a point $P \in M$ by an additional index characterizing the chart U_i: $x^k_i = x^k_i(P)$. Since the point P can belong simultaneously to several charts, it acquires several sets of local coordinates.

Here are simple examples of manifolds.

Example 1. In the Introduction we have considered a circle $S^1 \subset \mathbf{R}^2$ defined by the equation $x^2 + y^2 = 1$. Let us cover S^1 with an atlas consisting of four charts (Figs. 3.1 and 3.2)

$$U_1\ \{(x, y) \in S^1: y > 0\}, \quad U_2 = \{(x, y) \in S^1: y < 0\},$$
$$U_3 = \{(x, y) \in S^1: x > 0\}, \quad U_4 = \{(x, y) \in S^1: x < 0\}.$$

* Strictly speaking, the Cartesian coordinates $(x^1, \ldots, x^n)$ in a Euclidean space $\mathbf{R}^n$ are linear functions defined in $\mathbf{R}^n$, and the correspondence associating with a vector $\mathbf{a} \in \mathbf{R}^n$ a set of its coordinates $(x^1(\mathbf{a}), \ldots, x^n(\mathbf{a}))$ is, in fact, a mapping of $\mathbf{R}^n$ into an arithmetic linear n-dimensional space. It would therefore be better to write $x^k = x^k(P) = x^k(\varphi_i(P))$, but for the sake of brevity we shall omit the symbol φ_i if no confusion arises.

The corresponding domains V_1, V_2, V_3, and V_4 on the real axis $\mathbf{R}^1$ coincide and are equal to the open interval $(-1, 1)$. Homeomorphisms φ_1 and φ_2 are constructed as projections of the circle onto the x-axis: $\varphi_1(x, y) = \varphi_2(x, y) = x$, and homeomorphisms φ_3 and φ_4 as projections onto the y-axis: $\varphi_3(x, y) = \varphi_4(x, y) = y$. In order to prove that the mappings φ_k, $k = 1, \ldots, 4$, are homeomorphisms, it is sufficient to write explicitly the inverse mappings

$$\varphi_1^{-1}(x) = (x, \sqrt{1-x^2}) \in S^1, \quad \varphi_2^{-1}(x) = (x, -\sqrt{1-x^2}) \in S^1,$$

$$\varphi_3^{-1}(y) = (\sqrt{1-y^2}, y) \in S^1, \quad \varphi_4^{-1}(y) = (-\sqrt{1-y^2}, y) \in S^1$$

and demonstrate that they are continuous. Then we obtain on the circle four local coordinate systems, each consisting only of one coordinate: $x_1 = \varphi_1(x, y) = x$, $x_2 = \varphi_2(x, y) = x$, $x_3 = \varphi_3(x, y) = y$, $x_4 = \varphi_4(x, y) = y$. Certain points are provided with two local coordinate systems. For instance, for points P of the intersection $U_1 \cap U_3$ the coordinates $x_1(P)$ and $x_3(P)$ are valid (Fig. 3.3). There are other ways of introducing an atlas on a circle. In Chapter 1 we have considered polar coordinates (r, φ) on a plane. The equation of a circle in these coordinates is very simple: $r = 1$. Strictly speaking, polar coordinates on a plane are not a coordinate system (see Chapter 1). We introduce therefore two charts on a circle S^1, namely, $U_1 = \{(x, y) \in S^1 \colon x \neq -1\}$ and $U_2 = \{(x, y) \in S^1 \colon x \neq 1\}$ (Fig. 3.4). Let $\varphi_1(P) = \varphi_1(x, y)$ be the value of φ in the interval $(-\pi, \pi)$ and $\varphi_2(P) = \varphi_2(x, y)$ be the value of φ in the interval $(0, 2\pi)$, i.e. $V_1 = (-\pi, \pi)$, $V_2 = (0, 2\pi)$. Obviously, the local coordinates $\varphi_1 = \varphi_1(P)$ and $\varphi_2 = \varphi_2(P)$ coincide for points of the upper semicircle and do not coincide for points of the lower semicircle, that is, $\varphi_1(x, y) = \varphi_2(x, y)$ for $y > 0$ and $\varphi_1(x, y) = \varphi_2(x, y) - 2\pi$ for $y < 0$ (Fig. 3.5).

Example 2. The circle S^1 considered in Example 1 is a rather complicated manifold. The simplest example is represented by Euclidean space $\mathbf{R}^n$. We may take an atlas consisting of only one chart $U = \mathbf{R}^n$, the coordinate homeomorphism φ is the identity mapping $\varphi \colon U \to V = \mathbf{R}^n$, the local coordinate system is Cartesian coordinates of points in $\mathbf{R}^n$. Similarly, any domain $U \subset \mathbf{R}^n$ is an n-dimensional manifold whose atlas also consists of one chart with a Cartesian coordinate system.

Example 3. Let $f \colon \mathbf{R}^n \to \mathbf{R}^1$ be a continuous function and $\Gamma_f \subset R^{n+1}$ its *graph*, i.e. the set of points $(x_1, \ldots, x^n, x^{n+1})$: $x^{n+1} = f(x^1, \ldots, x^n)$. The space Γ_f is an n-dimensional manifold with an atlas consisting of one chart $U = \Gamma_f$. The coordinate homeomorphism $\varphi \colon U \to V = \mathbf{R}^n$ will be defined as a projection along the last coordinate: $\varphi(x^1, \ldots, x^n, x^{n+1}) = (x^1, \ldots, x^n) \in \mathbf{R}^n$. Then the inverse mapping φ^{-1} is given by $\varphi^{-1}(x^1, \ldots, x^n) = (x^1, \ldots, x^n, f(x^1, \ldots, x^n))$ and is, apparently, a continuous mapping.

Example 4. Let us consider an *n-dimensional* sphere S^n of unit radius defined as a set of points in $\mathbf{R}^{n+1}$ satisfying the equation $(x^1)^2 + (x^2)^2 + \ldots + (x^{n+1})^2 = 1$. We shall demonstrate that an n-dimensional sphere is an n-dimensional manifold. The open sets

$$U_i^+ = \{(x^1, \ldots, x^{n+1}) \in S^n: x^i > 0\},$$
$$U_i^- = \{(x^1, \ldots, x^{n+1}) \in S^n: x^i < 0\},$$

will be taken as an atlas. We obtain $2n + 2$ open sets covering the entire sphere S^n. Indeed, if the point $P = (x^1, \ldots, x^{n+1})$ belongs

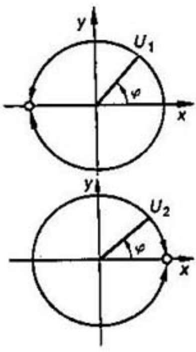

Figure 3.4

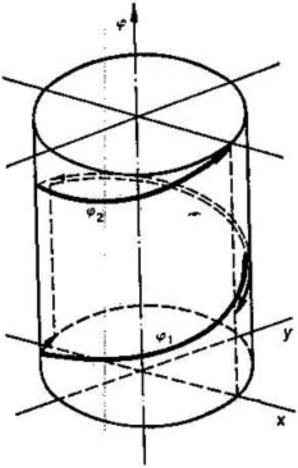

Figure 3.5

neither to charts U_i^+ nor to charts U_i^- for all i, then the inequalities $x^i \leqslant 0$ and $x^i \geqslant 0$ are satisfied, i.e. $x^i = 0$, $i = 1, 2, \ldots, n + 1$. Then $(x^1)^2 + \ldots + (x^{n+1})^2 = 0$, that is, the point P does not lie on the sphere S^n. The coordinate homeomorphisms φ_i^+ and φ_i^- are defined as projections of the Euclidean space $\mathbf{R}^{n+1}$ onto $\mathbf{R}^n$ along the coordinate x^i. In this case the domains V_i^+ and V_i^- coincide and are equal to a unit ball, and the coordinate homeomorphisms are given by

$$\varphi_i^+ (x^1, \ldots, x^{n+1}) = \varphi_i^- (x^1, \ldots, x^{n+1})$$
$$= (x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^{n+1}) \in \mathbf{R}^n.$$

The inverse homeomorphisms are defined by the formulas

$$(\varphi_i^+)^{-1}(y^1, \ldots, y^n)$$
$$= (y^1, \ldots, y^{i-1}, \sqrt{1-(y^1)^2-\ldots-(y^n)^2}, y^i, \ldots, y^n) \in \mathbf{R}^{n+1},$$
$$(\varphi_i^-)^{-1}(y^1, \ldots y^n)$$
$$= (y^1, \ldots, y^{i-1}, -\sqrt{1-(y^1)^2-\ldots-(y^n)^2}, y^i, \ldots, y^n) \in \mathbf{R}^{n+1},$$

and are apparently continuous.

Example 5. Let us consider a projective plane $\mathbf{RP}^2$. By this plane we mean a space the points of which are straight lines through the origin in $\mathbf{R}^3$. Define the distance between two straight lines as the least angle between them. Then $\mathbf{RP}^2$ becomes a metric space. We now prove that $\mathbf{RP}^2$ is a two-dimensional manifold. To this end, it is convenient to describe any straight line $P \in \mathbf{RP}^2$ by three homogeneous coordinates $(x : y : z)$ which admit multiplication by a number $\lambda \neq 0$. Homogeneous coordinates do not vanish simultaneously, i.e. $x^2 + y^2 + z^2 > 0$. Cover $\mathbf{RP}^2$ with three charts: $U_1 = \{(x : y : z)\colon x \neq 0\}$, $U_2 = \{(x : y : z)\colon y \neq 0\}$, and $U_3 = \{(x : y : z)\colon z \neq 0\}$. Let $V_1 = V_2 = V_3 = \mathbf{R}^2$. The mappings $\varphi_k\colon U_k \to V_k = \mathbf{R}^2$ are taken as coordinate homeomorphisms, namely, $\varphi_1(x : y : z) = (y/x, z/x)$, $\varphi_2(x : y : z) = (x/y, z/y)$, $\varphi_3(x : y : z) = (x/z, y/z)$. Thus, we have constructed three local coordinate systems

$$x_1^1 = y/x,\ x_1^2 = z/x;\ x_2^1 = x/y,\ x_2^2 = z/y;\ x_3^1 = x/z,\ x_3^2 = y/z.$$

To verify that the mappings φ_k are homeomorphisms, it is sufficient to construct the inverse mappings

$$\varphi_1^{-1}(x_1^1, x_1^2) = (1 : x_1^1 : x_1^2), \quad \varphi_2^{-1}(x_2^1, x_2^2) = (x_2^1 : 1 : x_2^2),$$
$$\varphi_3^{-1}(x_3^1, x_3^2) = (x_3^1 : x_3^2 : 1)$$

and prove that they are continuous.

The examples presented above show that the same manifold M admits distinct atlases. Even though the charts, as open sets, remain unchanged, we can alter the local coordinate system in a chart by choosing another coordinate homeomorphism. In particular, the following lemma holds true.

Lemma 1. *Let M be an n-dimensional manifold and let U be its chart with a coordinate homeomorphism φ and a local coordinate system $(x^1, \ldots, x^n)$. If $U' \subset U$ is an open subset of U, a coordinate homeomorphism φ' and a local coordinate system $(y^1, \ldots, y^n)$ can also be defined on U', and in this case $\varphi'(P) = \varphi(P)$, $y^k(P) = x^k(P)$ for $P \in U'$.*

Proof of Lemma 1 follows from the fact that the homeomorphism $\varphi\colon U \to V$ maps homeomorphically any open subset $U' \subset U$. It is

sufficient therefore to take the restriction of φ to U' as φ' and the restriction of coordinate functions x^h to the same subset U' as y^h.

Lemma 1 shows that using a given atlas $\{U_i\}$ we can construct a new atlas consisting of finer charts. On the other hand, the union of two atlases $\{U_i\}$ and $\{U'_j\}$ is again an atlas of the manifold. Thus, there exists a *maximal atlas* consisting of all the charts of a given manifold. A maximal atlas may be considered as a union of all atlases on a manifold.

Now we shall prove another useful lemma.

Lemma 2. *Let $\{U_i\}$ and $\{U'_j\}$ be two atlases on a manifold M. Then there exists a third atlas which refines these two atlases.*

To *prove* the lemma, we put $W_{ij} = U_i \cap U'_j$. According to Lemma 1, a local coordinate system can be introduced on each open set W_{ij}. On the other hand, $W_{ij} \subset U_i$ and $W_{ij} \subset U'_j$, so that the system of sets $\{W_{ij}\}$ covers M. Hence, $\{W_{ij}\}$ is an atlas refining both $\{U_i\}$ and $\{U'_j\}$.

3.1.2. FUNCTIONS OF COORDINATE TRANSFORMATION. DEFINITION OF A SMOOTH MANIFOLD

Among metric (or, in general, topological) spaces manifolds seem to be of special interest. For example, any continuous function $f\colon M \to \mathbf{R}^1$ defined on an n-dimensional manifold M in the neighbourhood of each point $P \in M$ can be identified with an ordinary continuous real-valued function $h(x^1, \ldots, x^n)$ of n independent real variables $(x^1, \ldots, x^n)$, the function h being defined in a domain of a Euclidean space $\mathbf{R}^n$. Indeed, let U be a chart containing a point P, let $\varphi\colon U \to V \subset \mathbf{R}^n$ be a coordinate homeomorphism of this chart, and let $(x^1(P), \ldots, x^n(P))$ be a local coordinate system in U. If $\mathbf{x} = (x^1, \ldots, x^n)$ is a vector with coordinates $(x^1, \ldots, x^n)$, we put $h(x^1, \ldots, x^n) = f(\varphi^{-1}(\mathbf{x}))$. Conversely, if h is a continuous function of n real variables defined in a domain $V \subset \mathbf{R}^n$, we can associate with h a continuous function f valid in the domain U of the manifold M: $f(P) = h(x^1(P), \ldots, x^n(P))$.

More generally, let $f\colon M_1 \to M_2$ be a continuous mapping of an n-dimensional manifold M_1 into an m-dimensional manifold M_2. Suppose $Q_0 = f(P_0)$, $P_0 \in M_1$ and $Q_0 \in M_2$. Then in a small neighbourhood $U \ni P_0$ the mapping f can be identified with a continuous vector function h of n independent variables. Indeed, let $U' \ni Q_0$ be a chart of the manifold M_2 and let $(y^1, \ldots, y^m)$ be a local coordinate system. Since the mapping f is continuous, there exists, according to Lemma 1, a chart U of point P_0 such that $f(U) \subset U'$. Suppose $(x^1, \ldots, x^n)$ is a local coordinate system in U. Since points P of the chart U are in a one-to-one correspondence with their coordinates $(x^1(P), \ldots, x^n(P))$, and points Q of the chart U' are also in a one-to-one correspondence with their coordinates $(y^1(Q), \ldots, y^m(Q))$.

the equality $Q = f(P)$ means

$$y^k(Q) = y^k(f(P)) = y^k(h(x^1(P), \ldots, x^n(P)) = h^k(x^1(P), \ldots, x^n(P)).$$

The functions $h^k(x^1, \ldots, x^n)$ are continuous, and the mapping f is uniquely reconstructed in U by these functions.

Thus, any continuous function f on a manifold M in a local coordinate system can be represented by a real-valued function h of n independent variables. If we alter the local coordinate system, the function h will also be modified. What is the law of modification of h under coordinate transformation? Let $(x_1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ be two local coordinate systems. Without loss of generality, we may assume that these coordinate systems are defined in the same chart U. Suppose h and h' are functions of coordinates $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$, respectively, which represent the function f. Then

$$f(P) = h(x^1(P), \ldots, x^n(P)) = h'(y^1(P), \ldots, y^n(P)). \qquad (1)$$

Since the coordinates $y^1, \ldots, y^n$ are also continuous functions in U, they can in turn be represented as functions of n independent variables $(x^1, \ldots, x^n)$, i.e.

$$\begin{aligned} y^1(P) &= y^1(x^1(P), \ldots x^n(P)), \\ &\vdots \\ y^n(P) &= y^n(x^1(P), \ldots, x^n(P)). \end{aligned} \qquad (2)$$

In these equations we deliberately use the same symbol y^h to denote both the coordinate of point P and its representation as a function of $(x^1, \ldots, x^n)$: $y^k = y^k(x^1, \ldots, x^n)$. Then from equation (1) we obtain the identity

$$h(x^1, \ldots, x^n) = h'(y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n)). \qquad (3)$$

The functions $y^k = y^k(x^1, \ldots, x^n)$ on the right-hand side of relations (2) were called in Chapter 1 functions of coordinate transformation, provided the set U is a domain in a Euclidean space. We shall retain this term for manifolds as well.

Definition 1. Let M be an n-dimensional manifold, $\{U_i\}$ its atlas, φ_i coordinate homeomorphisms, and $\{x_i^k\}$ a set of local coordinate systems. In each intersection of two charts $U_{ij} = U_i \cap U_j$ two local coordinate systems $\{x_i^k\}$ and $\{x_j^k\}$ are valid such that $x_i^k(P) = x_i^k(x_j^1(P), \ldots, x_j^n(P))$, $P \in U_{ij}$. The functions $x_i^k = x_i^k(x_j^1, \ldots, x_j^n)$ are called *functions of coordinate transformation* or *functions of transition* from the coordinates $\{x_i^k\}$ to the coordinates $\{x_j^k\}$.

Transition functions are not defined in the entire domain V_j, but only in its part $V_{ji} = \varphi_j(U_{ij})$ where it is meaningful to speak about two coordinate systems.

In Fig. 3.6 the domains V_i and V_j in a Euclidean space are shown, for convenience, as disjoint sets.

The transition functions $x_i^k = x_i^k(x_j^1, \ldots, x_j^n)$ map the domain V_{ji} into V_{ij} in $\mathbf{R}^n$

$$\mathrm{x}_i = \{x_i^k(x_j^1, \ldots, x_j^n)\} = \{x_j^1(\mathrm{x}_j)\} = \mathrm{x}_i(\mathrm{x}_j) = \varphi_i\varphi_j^{-1}(\mathrm{x}_j) = \varphi_{ij}(\mathrm{x}_j). \quad (4)$$

The mappings φ_{ij}: $V_{ji} \to V_{ij}$ given by equation (4) are, in fact, another writing of transition functions, and represent a homeomorphism of the domain V_{ji} onto V_{ij}.

Note that if $i = j$, then $U_{ij} = U_i$, $V_{ij} = V_{ji} = V_i$, and $x_i^k(x_j^1, \ldots, x_j^n) \equiv x_j^k$. Let us now return to the representation of a con-

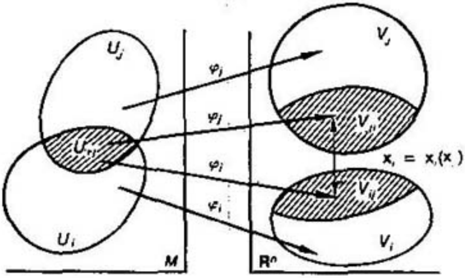

Figure 3.6

tinuous function f defined on an n-dimensional manifold M as a function h of n independent variables, local coordinates of a point of the manifold. It is known that a narrower class of function—differentiable functions—is of great significance in mathematical analysis. We now transfer this important concept to functions defined on a manifold. If a function $h(x^1, \ldots, x^n)$ is continuously differentiable, we cannot say the same about the function $h'(y^1, \ldots, y^n)$ representing f in another local coordinate system $(y^1, \ldots, y^n)$. Indeed, the functions h and h' are related by equation (3). Thus, the condition that h' is also continuously differentiable is that the functions of coordinate transformation $x^k = x^k(y^1, \ldots, y^n)$ should be continuously differentiable. If these functions are not continuously differentiable, there exists a function f such that its representation h in the coordinates $(x^1, \ldots, x^n)$ is a continuously differentiable function, while the representation h' in the coordinates $(y^1, \ldots, y^n)$ is not. As an example, we consider the function $f(P) = x^k(P)$, $P \in U \subset M$. Then $h(x^1, \ldots, x^n) \equiv x^k$ is, apparently, a continuously differentiable function, while $h'(y^1, \ldots, y^n) = x^k(y^1, \ldots, y^n)$ does not possess this property.

We thus arrive at the following definition.

Definition 2. A *smooth n-dimensional manifold* is an n-dimensional manifold M with an atlas $\{U_i\}$ having local coordinate systems $\{x_i^k\}$ satisfying the condition: the transition functions $x_i^k = x_i^k(x_j^1, \ldots, x_j^n)$ are continuously differentiable for any pair of charts U_i and U_j in the entire domain of definition.

This definition enables a class of continuously differentiable functions to be distinguished among all functions valid on a manifold M.

Definition 3. A function $f: M \to \mathbf{R}^1$ defined on a smooth manifold M is called *continuously differentiable* at a point $P_0 \in M$ if in any local coordinate system $(x_i^1, \ldots, x_i^n)$ (the charts $U_i \ni P_0$ belong to a fixed atlas) the function f can be represented as a continuously differentiable function $h(x_i^1, \ldots, x_i^n)$ of n independent variables in a neighbourhood of the point $(x_i^1(P_0), \ldots, x_i^n(P_0))$.

Note that the condition of continuous differentiability of the transition function in Definition 2 is essential for Definition 3. As we have already pointed out, if the transition functions were not continuously differentiable, the condition of continuous differentiability of f at point $P_0 \in M$ would depend on the choice of the chart U_i containing P_0.

Example 6. We now consider the following atlas on a manifold M. Let $M = \mathbf{R}^1$ be a real axis and let the atlas consist of two identical charts $U_1 = U_2 = M = \mathbf{R}^1$, but with distinct coordinate systems. On U_1 we define the coordinate $x_1 = x$, $x \in \mathbf{R}^1$, and on U_2 the coordinate $x_2 = x^3$. Then the transition functions are

$$x_2 = x_2(x_1) = (x_1)^3, \tag{5}$$

$$x_1 = x_1(x_2) = \sqrt[3]{x_2}. \tag{6}$$

While the transition function (5) is continuously differentiable (a polynomial), the function (6) has a discontinuous derivative. Thus, according to Definition 2, the manifold M with the atlas $\{U_1, U_2\}$ is not a smooth manifold.

Remark. If an atlas on a manifold M consists of only one chart (i.e. M is homeomorphic to a Euclidean domain), M is a smooth manifold.

Definition 4. Let on a manifold M there be given two atlases $\{U_i\}$ and $\{U'_j\}$ such that M is smooth with respect to each atlas. Two atlases $\{U_i\}$ and $\{U'_j\}$ are called *equivalent* if every function of coordinate transformation from any local coordinate system in $\{U_i\}$ to any local coordinate system in $\{U'_j\}$ is continuously differentiable.

A substantiation behind this definition lies in the fact that *any function f on a manifold M is continuously differentiable in the atlas $\{U_i\}$ if and only if it is continuously differentiable in $\{U'_j\}$*. Thus, from the point of view of continuously differentiable functions on

a manifold M equivalent atlases "have equal rights", so that any of the equivalent atlases can be used to represent a function as a continuously differentiable real-valued function of independent variables (coordinates of a point). Definition 4 admits another formulation: *two atlases* $\{U_i\}$ *and* $\{U'_j\}$ *are equivalent if* M *is a smooth manifold with respect to a new atlas equal to the union of the initial atlases,* $\{U_i\} \cup \{U'_j\}$.

We often have to deal with narrower classes of functions. Recall that a real-valued function $h(x^1, \ldots, x^n)$ is smooth of class C^r $(r = 1, 2, \ldots, \infty)$ in a neighbourhood of a point $(x_0^1, \ldots, x_0^n)$ if in this neighbourhood all partial derivatives of h up to order r exist and are continuous. For $r = \infty$ this means that the function h has continuous partial derivatives of any order. Consequently, we shall impose on atlases of a manifold M the conditions formulated in the following definition.

Definition 2′. A manifold M with a fixed atlas $\{U_i\}$ is called a *smooth manifold of class* C^r $(r = 1, 2, \ldots, \infty)$ (or *C^r-manifold*) if all the transition functions are smooth of class C^r at all points of the domain of their definition.

Definition 3′. Let M be a C^r-manifold and let f be a continuous function on this manifold. The function f is called *smooth of class* C^s, $s \leqslant r$ (or *C^s-function*) *in a neighbourhood of a point* $P_0 \in M$ if any representation of f as a function h of local coordinates $(x^1, \ldots, x^n)$ (from a fixed atlas) is a C^s-function in a neighbourhood of the point $(x^1(P_0), \ldots, x^n(P_0))$. The function f is smooth of class C^s if it is of class C^s in a neighbourhood of each point P_0 in the domain of definition.

Example 7. Let us modify Example 6 by choosing the coordinate $x_2 = x + x \cdot |x|$ in the second chart U_2. Then M is a C^1-manifold, but it is not C^2-manifold.

Remark. Below, if not stated otherwise, we shall consider only C^∞-manifolds and functions on these manifolds.

In examples 1-4 we have considered manifolds with such atlases that these manifolds are always of class C^∞.

Geometry may also deal with more strict conditions on atlases and their transition functions. For example, if all transition functions are real-analytic, i.e. in a neighbourhood of each point in their domain these functions can be expanded into convergent Taylor series, the manifold is called a *real-analytic manifold*. A real-analytic manifold is C^∞-manifold.

A more important class of manifolds is represented by complex-analytic manifolds. Let M be a $2n$-dimensional manifold. $\{U_j\}$ its atlas, and $\varphi_j\colon U_j \to V_j \subset \mathbf{R}^{2n}$ its coordinate homeomorphisms. Identify a $2n$-dimensional Euclidean space $\mathbf{R}^{2n}$ with an n-dimensional complex linear space $\mathbf{C}^n$, assuming that complex coordinates of

a point $(z^1, \ldots, z^n)$ give rise to $2n$ real coordinates $(x^1, \ldots, x^n, y^1, \ldots, y^n)$, $z^h = x^h + iy^h$. Then, $2n$ coordinate functions $x_j^1(P), \ldots, x_j^n(P), y_j^1(P), \ldots, y_j^n(P)$ in the chart U_j are transformed into n complex-valued functions $z_j^h(P) = x_j^h(P) + iy_j^h(P)$. The functions $z_j^h(P)$ are called complex coordinates of a point in the chart U_j. In the intersection of two charts $U_j \cap U_l$ the transition functions are given by

$$\begin{aligned} x_j^1 &= x_j^1(x_l^1, \ldots, x_l^n, y_l^1, \ldots, y_l^n), \\ &\vdots \\ x_j^n &= x_j^n(x_l^1, \ldots, x_l^n, y_l^1, \ldots, y_l^n), \\ y_j^1 &= y_j^1(x_l^1, \ldots, x_l^n, y_l^1, \ldots, y_l^n), \\ &\vdots \\ y_j^n &= y_j^n(x_l^1, \ldots, x_l^n, y_l^1, \ldots, y_l^n). \end{aligned} \tag{7}$$

These functions can be represented as complex-valued functions of n independent variables

$$\begin{aligned} z_j^1 &= z_j^1(z_l^1, \ldots, z_l^n), \\ &\vdots \\ z_j^n &= z_j^n(z_l^1, \ldots, z_l^n). \end{aligned} \tag{8}$$

Functions (8) are called transition functions or functions of transformation of complex coordinates.

A manifold M with a fixed atlas $\{U_j\}$ and local complex coordinate systems $(z_j^1, \ldots, z_j^n)$ is called a *complex-analytic manifold*, provided all transition functions (8) are complex-analytic, i.e. they can be expanded in convergent Taylor series of complex variables in a neighbourhood of each point in the corresponding domain of definition.

As an example of a manifold admitting a complex-analytic structure, we shall consider a two-dimensional sphere S^2 with a specially defined atlas. In Chapter 1 we have constructed the stereographic projection of the sphere $S^2 = x^2 + y^2 + z^2 = 1$ from the north pole $P_0 = (0, 0, 1)$ onto the coordinate plane (x, y). This projection, denoted by φ_0, maps all the points of S^2 except for the pole P_0 (i.e. the open set $U_0 = S^2 \setminus (P_0)$) homeomorphically onto the entire plane $V_0 = \mathbf{R}^2$. In Cartesian coordinates the homeomorphism φ_0 is of the form $\varphi_0(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$. We introduce therefore in the chart U_0 one complex coordinate $w_0 = \frac{x+iy}{1-z}$ expressed in terms of the Cartesian coordinates on a sphere. Furthermore, we shall consider the south pole $P_1 = (0, 0, -1)$ and the stereographic projection φ_1 from the south pole onto the same coordinate plane (x, y). The projection φ_1 maps homeomorphically the

set $U_1 = S^2 \setminus (P_1)$ onto the entire plane $V_1 = \mathbf{R}^2$. In Cartesian coordinates φ_1 takes the form $\varphi_1(x, y, z) = \left(\frac{x}{1+z}, \frac{y}{1+z}\right)$. Introduce in the chart U_1 a complex coordinate $w_1 = \frac{x-iy}{1+z}$. Then, in the intersection $U_0 \cap U_1$ we obtain $w_0 w_1 = \frac{x^2+y^2}{1-z^2} \equiv 1$. Hence,

$$w_0 = w_0(w_1) = \frac{1}{w_1}, \quad w_1 = w_1(w_0) = \frac{1}{w_0}. \tag{9}$$

Functions (9) are complex-valued, so that the sphere S^2 is a complex-analytic manifold. Each chart U_0 and U_1 covers the entire sphere S^2, except for one point, and is identified by virtue of the coordinate homeomorphisms φ_0 and φ_1 with the complex plane $\mathbf{C}^1 = \mathbf{R}^2$. Thus, the sphere S^2 is usually identified with the so-called completed complex plane obtained from $\mathbf{C}^1$ by adjoining another "infinitely distant" point.

An arbitrary smooth manifold need not necessarily be a complex-analytic one. For example, if its dimension is odd, the manifold is not, by trivial considerations, complex-analytic. Yet, there exist manifolds of even dimension which do not admit a complex-analytic structure either. For instance, as is demonstrated in Sec. 4.6 of Chapter 4, a projective plane is not a complex-analytic manifold.

3.1.3. SMOOTH MANIFOLDS. DIFFEOMORPHISM

Let M_1 and M_2 be smooth manifolds and let $f\colon M_1 \to M_2$ be a continuous mapping. As was already noted, in a neighbourhood of any point $P_0 \in M_1$ the mapping f can be represented as a vector function h, $y^k = h^k(x^1, \ldots, x^n)$, where $(x^1, \ldots, x^n)$ is a local coordinate system in the neighbourhood of $P_0 \in M_1$ and $(y^1, \ldots, y^m)$ is a local coordinate system in the neighbourhood of point $Q_0 = f(P_0) \in M_2$.

Definition 5. The mapping $f\colon M_1 \to M_2$ of smooth manifolds is called a *smooth mapping of class C^r* ($r = 1, 2, \ldots, \infty$) (or *C^r-mapping*) if for arbitrary local coordinate system $(x^1, \ldots, x^n)$ in the neighbourhood of any point $P_0 \in M_1$ and $(y^1, \ldots, y^m)$ in the neighbourhood of point $Q_0 = f(P_0) \in M_2$ the representation of f as a vector function $\mathbf{y} = (y^k) = (h^k(x^1, \ldots, x^n)) = h(\mathbf{x})$ is a vector function of class C^r.

Note that the definition of a C^r-manifold has a meaning if only the manifolds M_1 and M_2 are smooth of class not less than C^r.

Let $f\colon M_1 \to M_2$ be a homeomorphism of manifolds. If f is a C^r-mapping, the inverse mapping f^{-1} need not be a smooth mapping. Therefore, if the inverse mapping $f^{-1}\colon M_2 \to M_1$ is also a C^r-mapping the homemorphism f is called a *smooth homeomorphism of class C^r* or *C^r-diffeomorphism*. Diffeomorphisms of smooth manifolds play the

same role as homeomorphisms of topological spaces. If $f\colon M_1 \to M_2$ is a diffeomorphism, the manifolds M_1 and M_2 are called diffeomorphic. The set of all manifolds is subdivided into non-intersecting classes of pairwise diffeomorphic manifolds. Any general property of smooth manifolds, smooth functions or mappings on a manifold can be transferred to any other diffeomorphic manifold. We shall not therefore distinguish between diffeomorphic manifolds.

There are however such properties of manifolds that their "identity" for a pair of diffeomorphisms is not quite obvious. In particular, we have assigned to each manifold a numerical characteristic, the dimension. Do diffeomorphic manifolds have the same dimension?

Theorem 1. *Let $f\colon M_1 \to M_2$ be a C^r-homeomorphism ($r \geqslant 1$) of smooth manifolds. Then* $\dim M_1 = \dim M_2$.

Proof. Suppose $P_0 \in M_1$ is an arbitrary point, $Q_0 = f(P_0)$, and $g = f^{-1}$ is the inverse mapping. Choose local coordinates $(x^1, \ldots, x^n)$ in the neighbourhood U_0 of point P_0 and local coordinates $(y^1, \ldots, y^m)$ in the neighbourhood V_0 of point Q_0. Then the mappings f and g can be represented as vector functions $\mathbf{x} = h^{-1}(\mathbf{y})$ and $\mathbf{y} = h(\mathbf{x})$, with $h(h^{-1}(\mathbf{x})) \equiv x$ and $h^{-1}(h(\mathbf{y})) \equiv \mathbf{y}$. Consider the mappings h and h^{-1}. The mapping h consists of m functions $y^l = h^l(x^1, \ldots, x^n)$ of n independent variables $(x^1, \ldots, x^n)$. Compose the matrix of all partial derivatives of h^l

$$dh = \begin{pmatrix} \dfrac{\partial h^1}{\partial x^1} & \cdots & \dfrac{\partial h^1}{\partial x^n} \\ \vdots & & \vdots \\ \dfrac{\partial h^m}{\partial x^1} & \cdots & \dfrac{\partial h^m}{\partial x^n} \end{pmatrix}.$$

The matrix dh is a rectangular one of order $(m \times n)$, i.e. it has m rows and n columns. The matrix dh is called the *Jacobi matrix* of the mapping h.

Lemma 3. *Let $U_0 \subset \mathbf{R}^n$, $V_0 \subset \mathbf{R}^m$, and $W_0 \subset \mathbf{R}^k$ be open Euclidean domains, let $f\colon U_0 \to V_0$ and $g\colon V_0 \to W_0$ be continuously differentiable mappings, and let $h\colon U_0 \to W_0$ be the composition of f and g, i.e. $h(P) = g(f(P))$. Then the relation $dh(P) = dg(f(P)) \cdot df(P)$, $P \in U_0$, holds true for the mappings f, g, and h. In other words, the Jacobi matrix of the composition $h = g \circ f$ is the product of the Jacobi matrices of the mappings g and f.*

The *proof* of the lemma directly follows from the rule of differentiation of a composite function. Let $(x^1, \ldots, x^n)$, $(y^1, \ldots, y^m)$, and $(z^1, \ldots, z^k)$ be Cartesian coordinates in the domains U_0, V_0, and W_0. Then

$$y^i = f^i(x^1, \ldots x^n), \quad z^j = g^j(y^1, \ldots, y^m),$$

$$z^j = h^j(x^1, \ldots, x^n) = g^j(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)).$$

Differentiating the functions h^j with respect to x^i, we obtain

$$\frac{\partial h^j}{\partial x^i}(x^1, \ldots, x^n) = \sum_{l=1}^{m} \frac{\partial g^j}{\partial y^l}(f^1(x^1, \ldots, x^n), \ldots) \frac{\partial f^l}{\partial x^i}(x^1, \ldots, x^n). \tag{10}$$

This formula exactly coincides with the definition of a common element of the product of two matrices, i.e. $dh(x^1, \ldots, x^n) = dg(f^1(x^1, \ldots, x^n), \ldots) \cdot df(x^1, \ldots, x^n)$. If a point P has the coordinates $(x^1, \ldots, x^n)$, this equality can be written compactly in the form $dh(P) = dg(f(P)) \cdot df(P)$. The lemma is proved.

We now apply Lemma 3 to a pair of mappings h and h^{-1}. Let e_0 be the composition of the mappings h and h^{-1} and let e_1 be the composition of h^{-1} and h, i.e. $e_0(Q) = h(h^{-1}(Q))$, $Q \in V_0$, and $e_1(P) = h^{-1}(h(P))$, $P \in U_0$.

The mappings e_0: $V_0 \to V_0$ and e_1: $U_0 \to U_0$ are both identity mappings, so that their Jacobi matrices are identity matrices of order m and n, respectively. In particular, rank $de_0 = m$ and rank $de_1 = n$. On the other hand, by Lemma 3 we find

$$de_0(Q) = dh(h^{-1}(Q)) \cdot dh^{-1}(Q), \quad Q \in V_0.$$
$$de_1(P) = dh^{-1}(h(P)) \cdot dh(P), \quad P \in U_0.$$

It is known from linear algebra that the rank of the product of two matrices does not exceed the rank of each matrix. Since rank $dh \leqslant \min(m, n)$ and rank $dh^{-1} \leqslant \min(m, n)$ (the matrices dh and dh^{-1} are rectangular!), rank $de_0 \leqslant \min(m, n)$ and rank $de_1 \leqslant \min(m, n)$. Hence, $m \leqslant \min(m, n)$ and $n \leqslant \min(m, n)$ or $\max(m\ n) \leqslant \min(m, n)$, that is, $m = n$. Theorem 1 is proved.

The concept of dimension is valid not only for smooth manifolds, but for arbitrary manifolds as well. A question naturally arises: do the dimensions of homeomorphic manifolds coincide? The answer is yes, i.e. *if M_1 and M_2 are two homeomorphic manifolds, then* $\dim M_1 = \dim M_2$. This is a fundamental theorem of general topology, but it is beyond the scope of our course.

In conclusion, we shall prove several useful statements concerning the structure of an atlas. By definition, an atlas on a manifold M consists of open sets U_i homeomorphic to domains $V_i \subset \mathbf{R}^n$. If M is a smooth manifold, the coordinate homeomorphisms φ_i: $U_i \to V_i$ are also smooth. It is convenient sometimes to simplify the form of domains V_i in $\mathbf{R}^n$, though at the expense of an increased number of charts in the atlas.

Lemma 4. *In a smooth manifold M there exists an atlas $\{U_i\}$ such that any chart U_i is diffeomorphic to $\mathbf{R}^n$.*

Proof. First, we shall demonstrate that there exists an atlas such that any of its charts is diffeomorphic to an open ball of radius ε in $\mathbf{R}^n$. Let $P_0 \in M$ be an arbitrary point, $U_i \ni P_0$, and let φ_i: $U_i \to$

$V_i \subset \mathbf{R}_n$ be a coordinate homeomorphism, $Q_0 = \varphi_i(P_0)$. Since V_i is an open set in $\mathbf{R}^n$, we can find a number ε such that the ε-ball about Q_0 lies in V_i. Denote this ball by $O_\varepsilon(Q_0)$ and its inverse image $\varphi_i^{-1}(O_\varepsilon(Q_0))$ by W_P. The family of open sets $\{W_P\}$ is an atlas on M and each chart W_P is diffeomorphic to an open ball in $\mathbf{R}^n$. To complete the proof of the lemma, we shall demonstrate that an ε-ball is diffeomorphic to $\mathbf{R}^n$. It is sufficient to consider the case $\varepsilon = 1$. Let $(x^1, \ldots, x^n)$ be a point in a unit ball, i.e. $(x^1)^2 + \ldots + (x^n)^2 < 1$. Put

$$y^k = \frac{x^k}{\sqrt{1-(x^1)^2-\ldots-(x^n)^2}}, \tag{11}$$

$$x^k = \frac{y^k}{\sqrt{1+(y^1)^2+\ldots+(y^n)^2}}. \tag{12}$$

Functions (11) and (12) are smooth and map a ball of radius 1 into $\mathbf{R}^n$ and vice versa.

Lemma 5. *Let M be a smooth compact manifold equipped with an atlas $\{U_i\}$. Then there exists a smooth partition of unity ψ_i subordinate to the covering $\{U_i\}$.*

Proof. According to Lemma 4, it suffices to assume that all charts are homeomorphic to a ball of radius 1. Let $\varphi_i\colon U_i \to D_1^n \subset \mathbf{R}^n$ be coordinate homeomorphisms.* Choose a rather small $\varepsilon > 0$ such that $\varphi_i^{-1}(D^n_{(1-\varepsilon)})$ cover the manifold M. Suppose there exists on the ball D_1^n a C^∞-function f such that $\operatorname{supp} f = D^n_{(1-\varepsilon)}$, $0 \leqslant f \leqslant 1$. Put

$$\bar\psi_i(P) = \begin{cases} 0 & \text{if } P \notin U_i, \\ f(\varphi_i(P)) & \text{if } P \in U_i. \end{cases}$$

Since $f(\varphi_i(P)) = 0$ for $P \notin \varphi_i^{-1} D^n_{(1-\varepsilon)}$, the functions $\bar\psi_i$ are smooth on M and $\operatorname{supp} \bar\psi_i \subset U_i$, $0 \leqslant \bar\psi_i \leqslant 1$. Moreover, $\operatorname{supp} \psi_i \supset \varphi_i^{-1} D^n_{(1-\varepsilon)}$, hence the sum $\bar\psi_i \bar\psi(P) = \sum\limits_i \bar\psi_i(P)$ is strictly positive at each point. We take then $\psi_i(P) = \bar\psi_i(P)/\bar\psi(P)$. The functions $\psi_i(P)$ form a smooth partition of unity subordinate to the covering $\{U_i\}$.

It remains therefore to construct a C^∞-function f in $\mathbf{R}^n$ such that its support is equal to the ball $D^n_{(1-\varepsilon)}$. We shall seek f in the form $f(x^1, \ldots, x^n) = h((x^1)^2 + \ldots + (x^n)^2)$. Hence, it is sufficient to find a smooth function of one variable $h(x)$ such that $h(x) = 0$ for $x > (1-\varepsilon)^2$ and $h(x) > 0$ for $x < (1-\varepsilon)^2$. We shall choose the function

$$h(x) = \begin{cases} e^{-1/(x-(1-\varepsilon)^2)^2}, & x < (1-\varepsilon)^2, \\ 0, & x \geqslant (1-\varepsilon)^2. \end{cases}$$

which is known to be smooth of class C^∞. Lemma 5 is proved.

* Here D_r^n is an r-ball about the origin.

Problems

1. Prove that the space of positions of a rigid segment on a plane is a smooth manifold.
2. Prove that the group $SO(3)$ is homeomorphic to a three-dimensional projective space.
3. Describe the configurational space of a system of two hinged bars in a three-dimensional space.
4. Give an example of a smooth one-to-one mapping which is not a diffeomorphism.
5. Demonstrate that no atlas consisting of one chart exists on a sphere S^n.

3.2. DEFINITION OF MANIFOLDS BY EQUATIONS

There is a conventional way, used most frequently in practice, to describe and construct manifolds.

In the preceding section (as well as in Chapter 1) many examples of manifolds appeared as a set of solutions of a non-linear equation given in a Euclidean space. For instance, an n-dimensional sphere S^n in a Euclidean space $\mathbf{R}^{n+1}$ is defined by the equation $(x^1)^2 + (x^2)^2 + \ldots + (x^{n+1})^2 = 1$; a pseudosphere S_1^2 is given by $x^2 + y^2 - z^2 = -1$. In general, if $f(x^1, \ldots, x^n)$ is a continuously differentiable function, the set of solutions of the equation $f(x^1, \ldots, x^n) - c = 0$ is called a *manifold of level c of the function f*. Thus, the Euclidean space $\mathbf{R}^n$ is decomposed into a union of level manifolds for the function f. In the case of a function of two variables, the solutions of the equation are called *level lines for f* and in the case of a function of three variables *level surfaces*.

To justify the term the level manifold for a function f, one should prove that the level manifold for f is really a manifold. However, this is not always the case.

Example 1. Let us consider the function $f(x, y) = x^2 - y^2$. Its level lines are described by the equation $x^2 - y^2 = c$. If $c > 0$, the level lines consist of two connectedness components, each being described by the equation $x = \sqrt{c + y^2}$ or $x = -\sqrt{c + y^2}$, i.e. these lines are the graphs of functions of one variable (Fig. 3.7). Similarly, for $c < 0$ the level lines are the graphs of two functions: $y = \sqrt{x^2 - c}$ and $y = -\sqrt{x^2 - c}$. Thus, for $c \neq 0$ the level lines are one-dimensional manifolds (see Sec. 3.1, Example 3). Of special interest is the level line for $c = 0$; it consists of two intersecting straight lines $y = x$ and $y = -x$, and is not a manifold. Indeed, we shall demonstrate that the point $P_0 = (0, 0)$ on the level line $c = 0$ does not possess a neighbourhood U_0 homeomorphic to a Euclidean domain. If such a homeomorphism $\varphi_0\colon U_0 \to V_0 \subset \mathbf{R}^1$ existed, we could assume, without loss of generality, that V_0 is an open

interval and U_0 contains all points P lying from P_0 at a distance less than ε. Then $U_0 \setminus (P_0)$ has at least four connectedness components, while its homeomorphic image $V_0 \setminus (\varphi_0(P_0))$ has only two connectedness components. This contradiction proves the absence of homeomorphism φ_0 for $\mathbf{R}^1$. In a similar way we can verify the absence of homeomorphism φ_0 for the domain $V_0 \subset \mathbf{R}^n$ for $n \geqslant 2$.

Nevertheless, a level manifold for a continuously differentiable function f is almost always a manifold.

Theorem 1. *Let $f = f(x^1, \ldots, x^n)$ be a function of class C^∞ defined in the entire Euclidean space $\mathbf{R}^n$ and let $M_c = \{(x^1, \ldots, x^n): f(x^1, \ldots, x^n) = c\}$. If the gradient of f is non-zero at each point of*

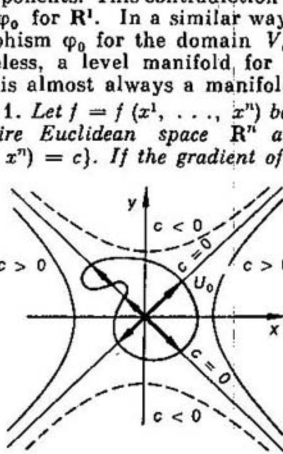

Figure 3.7

the set M_c, this set is a smooth $(n-1)$-dimensional manifold of class C^∞, and we can choose $(n-1)$ Cartesian coordinates of the ambient Euclidean space $\mathbf{R}^n$ as local coordinates in a neighbourhood of every point $P_0 \in M_c$.

Proof. The theorem is, in fact, the implicit function theorem formulated in convenient terms. Fix a point $P_0 \in M_c$, $P_0 = (x^1, \ldots, x_0^n)$. Since

$$\operatorname{grad}_{P_0} f \neq 0, \quad \operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}\right),$$

there exists a non-zero partial dervative at point P_0. Without loss of generality, we may assume $\frac{\partial f}{\partial x^n}(x_0^1, \ldots, x_0^n) \neq 0$. Let $Q_0 = (x_0^1, \ldots, x_0^{n-1})$ be a point in $\mathbf{R}^{n-1}$ which is the image of P_0 under projection along the coordinate axis x^n. According to the implicit function theorem, there exists a neighbourhood $V_0 \ni Q_0$ of point Q_0, an interval $(x_0^n - \delta, x_0^n + \delta)$, and a continuous function $y = y(x^1, \ldots, x^{n-1})$ of class C^∞ defined in V_0 such that:

(a) $f(x^1, \ldots, x^{n-1}, y(x^1, \ldots, x^{n-1})) \equiv c$ in V_0,

(b) $x_0^n = y(x_0^1, \ldots, x_0^{n-1})$,

(c) $|x_0^n - y(x^1, \ldots, x^{n-1})| < \delta$ in the domain V_0,

(d) any solution $(x^1, \ldots, x^n) \in V_0 \times (x_0^n - \delta, x_0^n + \delta)$ of the equation $f(x^1, \ldots, x^n) = c$ is of the form $x^n = y(x^1, \ldots, x^{n-1})$. Let $U_0 = M_c \cap (V_0 \times (x_0^n - \delta, x_0^n + \delta))$ stand for the neighbourhood of point $P_0 \in M_c$. This neighbourhood is the chart in question which contains P_0. Take the restriction of the projection $\mathbf{R}^n \to \mathbf{R}^{n-1}$ to U_0 as a coordinate homeomorphism $\varphi_0(x^1, \ldots, x^n) = (x^1, \ldots, x^{n-1}) \in V_0$ and define the inverse mapping φ_0^{-1} by the relation

$$\varphi_0^{-1}(x^1, \ldots, x^{n-1}) = (x^1, \ldots, x^{n-1}, y(x^1, \ldots, x^{n-1})).$$

It follows from condition (c) that $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in V_0 \times (x_0^n - \delta, x_0^n + \delta)$ and from condition (a) that $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in M_c$. Thus, $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in U_0$, i.e. the mappings φ_0 and φ_0^{-1} are continuous and mutually inverse.

We have proved that M_c is an $(n-1)$-dimensional manifold and found in the neighbourhood of each point $P_0 \in M_c$ a local coordinate system formed by some Cartesian coordinates in the Euclidean space $\mathbf{R}^n$. We now prove that the transition functions are smooth. Let P_0 be also contained in another chart U_1 and let Cartesian coordinates $(x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n)$ be chosen as local coordinates in U_1. Then in the intersection $U_0 \cap U_1$ the coordinates $(x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n)$ are expressed in terms of $(x^1, \ldots, x^{n-1})$ as

$$\begin{aligned} x^1 &= x^1, \\ &\vdots \\ x^{i-1} &= x^{i-1}, \\ x^{i+1} &= x^{i+1}, \\ &\vdots \\ x^{n-1} &= x^{n-1}, \\ x^n &= y(x^1, \ldots, x^{n-1}). \end{aligned} \tag{1}$$

Since the function $y = y(x^1, \ldots, x^{n-1})$ is smooth of class C^∞, all the functions (1) are also of class C^∞. This completes the proof of Theorem 1.

Example 2. Consider an n-dimensional sphere S^n given by the equation $f(x^1, \ldots, x^{n+1}) = \sum_{k=1}^{n+1} (x^k)^2 = 1$. The gradient of f is $\operatorname{grad} f = (2x^1, 2x^2, \ldots, 2x^{n+1})$. If the point $P = (x^1, \ldots, x^{n+1})$ lies on S^n, then not all of its coordinates vanish, that is, one of the gradient coordinates is non-zero. Conditions of Theorem 1 are satisfied, hence S^n is a C^∞-manifold.

Example 3. Consider a Euclidean space $\mathbf{R}^{n^2}$ of dimension n^2 and represent the points of $\mathbf{R}^{n^2}$ as square matrices A of order n with the coordinates $A = (a_{ij})$. Consider also the set $SL(n, \mathbf{R})$ of all matrices $A \in \mathbf{R}^{n^2}$ with the determinant equal to unity, $\det A = 1$. The set $SL(n, \mathbf{R})$ is a group with respect to multiplication of matrices and is called a special linear group. We shall demonstrate that $SL(n, \mathbf{R})$ is a C^∞-manifold of dimension $n^2 - 1$. Consider a function f of n^2 variables, $f(a_{ij}) = \det(a_{ij})$. This function is a polynomial and is therefore of class C^∞. In order to apply Theorem 1, we have to calculate grad f at all points of the group $SL(n, \mathbf{R})$. Let E be an identity matrix. Since $\det E = 1$, $E \in SL(n, \mathbf{R})$. Calculate grad f at point E. To this end, we express $\det A$ as

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \ldots + (-1)^{n+1} \det A_{1n} \cdot a_{1n}. \quad (2)$$

This expression contains the determinants of the matrices A_{1k} which are polynomials of all variables (a_{ij}) except those in the first row. Then the partial derivative of f with respect to a_{11} is of the form $\frac{\partial f}{\partial a_{11}} = \frac{\partial}{\partial a_{11}} (a_{11} \det A_{11}) = \det A_{11}$. At point E we obtain

$$\frac{\partial f}{\partial a_{11}} (E) = 1. \quad (3)$$

Thus, the gradient of f at point E is non-zero. We now demonstrate that at an arbitrary point $A_0 \in SL(n, \mathbf{R})$ grad f is also non-zero. Introduce new variables b_{ij} defined by $(b_{ij}) = B = A_0^{-1}A = A_0^{-1} \cdot (a_{ij})$. If $A = A_0$, then $B = E$ and

$$f(A) = f(A_0B) = \det(A_0B) = \det A_0 \cdot \det B = \det A_0 \cdot f(B).$$

Differentiating the superposition of functions, we obtain

$$\frac{\partial f}{\partial b_{11}} (E) = \sum_{ij} \frac{\partial f}{\partial a_{ij}} (A_0) \cdot \frac{\partial a_{ij}}{\partial b_{11}}. \quad (4)$$

The left-hand side of Eq. (4) is equal to unity according to formula (3); hence, at least one of the terms on the right-hand side of Eq. (4) is non-zero and therefore one of the partial derivatives $\frac{\partial f}{\partial a_{ij}} (A_0)$, as well as grad f, is also non-zero. Thus, the conditions of Theorem 1 are satisfied, so that the group $SL(n, \mathbf{R})$ is a smooth manifold of dimension $(n^2 - 1)$.

Theorem 1 can easily be extended to a system of non-linear equations. Note that Theorem 1 admits another formulation. The gradient of a function f can be represented as a column of partial derivatives of f and is therefore the Jacobi matrix of f. Then the condition of nontriviality of grad f at a point $P_0 \in \mathbf{R}^n$ is equivalent to the condition that the rank of the Jacobi matrix df of the function f is equal to unity, i.e. is maximal.

Let there be given a system of equations

$$\begin{aligned} f^1(x^1, \ldots, x^n) &= c^1, \\ f^2(x^1, \ldots, x^n) &= c^2, \\ &\vdots \\ f^k(x^1, \ldots, x^n) &= c^k, \end{aligned} \tag{4'}$$

which can be written in compact form as $f(\mathbf{x}) = \mathbf{c}$, where $\mathbf{x} = (x^1, \ldots, x^n) \in \mathbf{R}^n$, $\mathbf{c} = (c^1, \ldots, c^k) \in \mathbf{R}^k$, and f is a mapping defined by the functions $(f^1, \ldots, f^k)$. The set $M_\mathbf{c}$ of solutions of system (4') is called a level manifold for the system of functions $(f^1, \ldots, f^k)$.

Theorem 2. *Let $f\colon \mathbf{R}^n \to \mathbf{R}^k$ be a C^∞-mapping and let $M_\mathbf{c}$ be a set of solutions of the system of equations $f(\mathbf{x}) = \mathbf{c}$. If the rank of the Jacobi matrix of f is maximal at every point $P_0 \in M_\mathbf{c}$ (i.e. rank $df(P_0) = k$), $M_\mathbf{c}$ is an $(n-k)$-dimensional smooth manifold of class C^∞, and $(n-k)$ Cartesian coordinates of the surrounding Euclidean space $\mathbf{R}^n$ can be taken as local coordinates in the neighbourhood of each point $P_0 \in M_\mathbf{c}$.*

The *proof* of Theorem 2 is exactly analogous to that of Theorem 1 with the only difference that instead of one variable x^n we choose k variables $(x^{i_1}, \ldots, x^{i_k})$. Denoting this group of variables by one symbol, say $\mathbf{y} = (x^{i_1}, \ldots, x^{i_k})$, we obtain the same formulas as in Theorem 1.

Example 4. Let us consider in a Euclidean space $\mathbf{R}^4$ with the coordinates (x^1, x^2, x^3, x^4) the system of equations

$$(x^1)^2 + (x^2)^2 = 1, \quad (x^3)^2 + (x^4)^2 = 1. \tag{5}$$

The corresponding functions f^1 and f^2 are of the form

$$f^1(x^1, x^2, x^3, x^4) = (x^1)^2 + (x^2)^2, \; f^2(x^1, x^2, x^3, x^4) = (x^3)^2 + (x^4)^2.$$

In order to apply Theorem 2, we shall calculate the Jacobi matrix of the mapping $f = (f^1, f^2)$

$$df = \begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \frac{\partial f^1}{\partial x^2} & \frac{\partial f^1}{\partial x^3} & \frac{\partial f^1}{\partial x^4} \\ \frac{\partial f^2}{\partial x^1} & \frac{\partial f^2}{\partial x^2} & \frac{\partial f^2}{\partial x^3} & \frac{\partial f^2}{\partial x^4} \end{pmatrix} = \begin{pmatrix} 2x^1 & 2x^2 & 0 & 0 \\ 0 & 0 & 2x^3 & 2x^4 \end{pmatrix}.$$

Clearly, rank $df \leqslant 1$ if only one of the rows of the Jacobi matrix is zero, which is impossible at the points representing solutions of system (5). Thus, the solutions of this system form a two-dimensional manifold of class C^∞. Since system (5) splits into two equations, each for its own group of variables, the set of the solutions of this

system can also be expressed as the Cartesian product of the solutions of each equation, i.e. the solutions of system (5) are represented as the product of two copies of a circle. This manifold is called a (two-dimensional) *torus*.

3.3. TANGENT VECTORS. TANGENT SPACE

In Chapter 1 we have seen that the so-called infinitesimal properties of space are of great help in the study of metric properties of curves and surfaces and, generally, of metric properties of domains in a Euclidean space. These are the properties defined in a very small neighbourhood of a fixed point P by neglecting small quantities of an order higher than the distance from P. In mathematical analysis we use a similar procedure of neglecting infinitesimal quantities while studying the behaviour of a function in a neighbourhood of a point. In the case of smooth manifolds there is also a natural desire to neglect infinitesimal quantities. One of such methods is to introduce special concepts analogous to tangent vector to a curve and tangent plane to a surface.

3.3.1 SIMPLE EXAMPLES

Let us consider a smooth curve $\mathbf{x} = \mathbf{x}(t) = (x^1(t), x^2(t), x^3(t))$ in a three-dimensional space $\mathbf{R}^3$, where t is a parameter. Fix a value t_0 and expand the vector function $\mathbf{x} = \mathbf{x}(t)$ in a Taylor series about the point t_0

$$\mathbf{x}(t_0 + \Delta t) = \mathbf{x}(t_0) + \frac{d\mathbf{x}}{dt}(t_0)\,\Delta t + O(\Delta t^2). \tag{1}$$

The first two terms on the right-hand side of Eq. (1) may be considered, on the one hand, as an approximation of $\mathbf{x}(t)$ in a neighbourhood of the point t_0 by a linear vector function. On the other hand, this linear function $\mathbf{y}(\Delta t) = \mathbf{x}(t_0) + \frac{d\mathbf{x}(t_0)}{dt}\cdot(\Delta t)$ defines in $\mathbf{R}^3$ a straight line through the point $P_0 = \mathbf{x}(t_0)$. Furthermore, among all straight lines through P_0 the straight line $\mathbf{y}(t)$ "approaches most closely" the initial curve $\mathbf{x}(t)$. Let us first explain the term: a straight line "approaches closely" a curve $\mathbf{x}(t)$. We say that the straight line $\mathbf{y}(t) = \mathbf{a} + \mathbf{b}t$ ($|\mathbf{b}| = 1$) *is tangent* to a curve $\mathbf{x}(t)$ at point $\mathbf{x}(t_0)$ if the distance from the point $\mathbf{x}(t)$ to the straight line is an infinitesimal quantity in comparison with the distance from $P_0 = \mathbf{x}(t_0)$ to $\mathbf{x}(t)$. The point P_0 lies then on the straight line $\mathbf{y}(t)$. We may assume $\mathbf{x}(t_0) = \mathbf{y}(t_0)$, so that $\mathbf{y}(t_0 + \Delta t) = \mathbf{x}(t_0) + \mathbf{b}\Delta t$. The distance from the point $\mathbf{x}(t_0 + \Delta t)$ to $\mathbf{y}(t)$ is

$$\begin{aligned}|\mathbf{x}(t_0 + \Delta t) - \mathbf{x}(t_0) - \mathbf{b}(\mathbf{x}(t_0 + \Delta t) - \mathbf{x}(t_0), \mathbf{b})| \\ = O(|\mathbf{x}(t_0 + \Delta t) - \mathbf{x}(t_0)|).\end{aligned} \tag{2}$$

Suppose $\frac{d\mathbf{x}}{dt}(t_0) \neq 0$. Expanding $\mathbf{x}(t)$ by formula (1), we obtain instead of Eq. (2)

$$\left| \frac{d\mathbf{x}}{dt}(t_0)\,\Delta t - \mathbf{b}\left(\frac{d\mathbf{x}}{dt}(t_0)\,\Delta t,\ \mathbf{b}\right) \right| = O(\Delta t^2)$$

for $\Delta t \to 0$ or, dividing by Δt,

$$\left| \frac{d\mathbf{x}}{dt}(t_0) - \mathbf{b}\left(\frac{d\mathbf{x}}{dt}(t_0),\ \mathbf{b}\right) \right| = O(\Delta t). \tag{3}$$

Since the left-hand side of Eq. (3) does not depend on Δt, we obtain by a limiting process for $\Delta t \to 0$

$$\frac{d\mathbf{x}}{dt}(t_0) = \mathbf{b}\left(\frac{d\mathbf{x}}{dt}(t_0),\ \mathbf{b}\right), \tag{4}$$

which means that the vectors $\mathbf{b}$ and $\frac{d\mathbf{x}}{dt}(t_0)$ are collinear. Thus, the linear part of the Taylor formula (1) for the vector function $\mathbf{x}(t)$

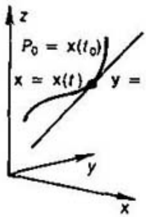

Figure 3.8

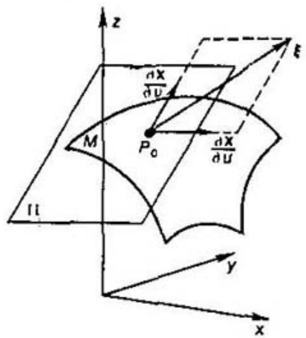

Figure 3.9

defines a parametric representation of a tangent straight line at point P_0 (Fig. 3.8). Consider now a surface M in a three-dimensional space $\mathbf{R}^3$ given in parametric form as a vector function $\mathbf{x} = \mathbf{x}(u, v)$ of two independent parameters u and v. The surface $\mathbf{x}(u, v)$ is called *non-degenerate* if at each point the partial derivatives $\frac{\partial \mathbf{x}}{\partial u}(u, v)$ and $\frac{\partial \mathbf{x}}{\partial v}(u, v)$ are linearly independent as vectors in $\mathbf{R}^3$. Fix parameters (u_0, v_0) and a plane Π, $\mathbf{x} = \mathbf{x}(u_0, v_0) + \mathbf{a}u + \mathbf{b}v$, through point $P_0 = \mathbf{x}(u_0, v_0)$ on the surface. The plane Π is called *tangent* to the surface M at point P_0 if the distance from Π to the point $\mathbf{x}(u, v)$

is an infinitesimal quantity in comparison with the distance from $\mathbf{x}(u, v)$ to P_0. Expanding the function $\mathbf{x}(u, v)$ in a Taylor series about the point (u_0, v_0)

$$\mathbf{x}(u_0+\Delta u,\ v_0+\Delta v)=\mathbf{x}(u_0,\ v_0)+\frac{\partial \mathbf{x}}{\partial u}(u_0,\ v_0)\,\Delta u + \frac{\partial \mathbf{x}}{\partial v}(u_0,\ v_0)\,\Delta v+O(\Delta u^2+\Delta v^2), \quad (5)$$

we find that the linear part in (5) defines a two-parameter representation of the tangent plane Π to the surface M at point $P_0 = \mathbf{x}(u_0, v_0)$ (Fig. 3.9). It is natural to call any vector emerging from P_0 and lying in the plane Π a *tangent vector* to the surface M at point P_0. It can be seen from Eq. (5) that parametric representation of a plane Π tangent to a surface M at a point P_0 is of the form

$$\mathbf{x}(\Delta u,\ \Delta v)=\mathbf{x}(u_0,\ v_0)+\frac{\partial \mathbf{x}}{\partial u}(u_0,\ v_0)\,\Delta u+\frac{\partial \mathbf{x}}{\partial v}(u_0,\ v_0)\,\Delta v. \quad (6)$$

Hence, any tangent vector $\boldsymbol{\xi}$ can be decomposed into a linear combination of the vectors $\frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)$ and $\frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)$

$$\boldsymbol{\xi}=\frac{\partial \mathbf{x}}{\partial u}(u_0,\ v_0)\,\Delta u+\frac{\partial \mathbf{x}}{\partial v}(u_0,\ v_0)\,\Delta v \quad (7)$$

for an appropriate choice of the parameters Δu and Δv. Thus, the vectors $\frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)$ and $\frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)$ form a basis in the tangent plane Π, and Δu and Δv are linear coordinates of the tangent vector $\boldsymbol{\xi}$ in this basis.

Let us draw a smooth curve $\mathbf{x} = \mathbf{x}(t)$ through point P_0 on a surface M. Since the curve $\mathbf{x} = \mathbf{x}(t)$ lies on the surface M, it can be represented parametrically as the composition

$$\mathbf{x}(t) = \mathbf{x}(u(t),\ v(t)) \quad (8)$$

of functions $u(t)$ and $v(t)$. In other words, the functions $u(t)$ and $v(t)$ define parametrically a curve in a local coordinate system (u, v) on the surface M. Then, the condition that the curve passes through point P_0 can be expressed as the condition for the coordinates: $u_0 = u(t_0)$, $v_0 = v(t_0)$. Calculation of a tangent vector to a curve (or, as it is frequently called, the velocity vector of a curve) yields

$$\begin{aligned}\frac{d\mathbf{x}}{dt}(t_0) &= \frac{d}{dt}(\mathbf{x}(u(t),\ v(t)))|_{t=t_0}\\ &= \frac{\partial \mathbf{x}}{\partial u}(u(t_0),\ v(t_0))\,\frac{du}{dt}(t_0)+\frac{\partial \mathbf{x}}{\partial v}(u(t_0),\ v(t_0))\,\frac{dv}{dt}(t_0)\\ &= \frac{\partial \mathbf{x}}{\partial u}(u_0,\ v_0)\,\frac{du}{dt}(t_0)+\frac{\partial \mathbf{x}}{\partial v}(u_0,\ v_0)\,\frac{dv}{dt}(t_0).\end{aligned}$$

Hence, *a tangent vector to a curve on a surface M lies in the tangent plane.*

Definition. Let $\xi = \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\xi^1 + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\xi^2$ be a tangent vector to a surface M at a point P_0. Then the numbers (ξ^1, ξ^2) are called *coordinates of the tangent vector ξ to M at point P_0 in a local coordinate system (u, v) on the surface M.*

This definition works not only for the coordinates (u, v) describing parametrically the surface M, but also for any coordinate system (u', v') in a neighbourhood of point P_0. Indeed, if (u', v') is some other coordinate system, the coordinates u and v are expressed as smooth functions of u' and v': $u = u(u', v')$, $v = v(u', v')$, $u_0 = u(u'_0, v'_0)$, $v_0 = v(u'_0, v'_0)$. Considering then compositions of the functions, we obtain a new parametric definition of a surface M

$$\mathbf{x} = \mathbf{x}(u(u', v'), v(u', v')). \tag{9}$$

The parametric equation of a tangent plane Π at point P_0 for the new parameters (u', v') is

$$\begin{aligned}
\mathbf{x} &= \mathbf{x}(u(u'_0, v'_0), v(u'_0, v'_0)) + \frac{\partial \mathbf{x}}{\partial u'_0}\Delta u' + \frac{\partial \mathbf{x}}{\partial v'_0}\Delta v' \\
&= \mathbf{x}(u_0, v_0) + \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\frac{\partial u}{\partial u'}\Delta u' + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\frac{\partial v}{\partial u'}\Delta u' \\
&\quad + \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\frac{\partial u}{\partial v'}\Delta v' + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\frac{\partial v}{\partial v'}\Delta v' \\
&= \mathbf{x}(u_0, v_0) + \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\left(\frac{\partial u}{\partial u'}(u'_0, v'_0)\Delta u' + \frac{\partial u}{\partial v'}(u'_0, v'_0)\Delta v'\right) \\
&\quad + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\left(\frac{\partial v}{\partial u'}(u'_0, v'_0)\Delta u' + \frac{\partial v}{\partial v'}(u'_0, v'_0)\Delta v'\right).
\end{aligned}$$

Assuming

$$\begin{aligned}
\Delta u &= \frac{\partial u}{\partial u'}(u'_0, v'_0)\Delta u' + \frac{\partial u}{\partial v'}(u'_0, v'_0)\Delta v', \\
\Delta v &= \frac{\partial v}{\partial u'}(u'_0, v'_0)\Delta u' + \frac{\partial v}{\partial v'}(u'_0, v'_0)\Delta v',
\end{aligned} \tag{10}$$

we arrive at the parametric definition of a tangent plane for the initial parameters (u, v)

$$\mathbf{x} = \mathbf{x}(u_0, v_0) + \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\Delta u + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\Delta v.$$

Thus, formula (9) is another parametric definition of the surface M with the same tangent plane Π. Curve (8) can therefore be written in terms of the functions $u'(t)$ and $v'(t)$ in such a way that $\mathbf{x}(t) = \mathbf{x}(u(u'(t), v'(t)), v(u'(t), v'(t)))$. Then, according to the definition of the coordinates of a tangent vector to a curve in a local

coordinate system (u', v'), the numbers $\left(\frac{du'}{dt}(t_0), \frac{dv'}{dt}(t_0)\right)$ are the coordinates of a tangent vector to a curve. Differentiating the composite functions u and v, we obtain a relation between the coordinates of a tangent vector to a curve in different local coordinate systems

$$\begin{aligned}\frac{\partial u}{\partial t}(t_0) &= \frac{\partial u}{\partial u'}(u'_0, v'_0)\frac{du'}{dt}(t_0) + \frac{\partial u}{\partial v'}(u'_0, v'_0)\frac{dv'}{dt}(t_0),\\ \frac{dv}{dt}(t_0) &= \frac{\partial v}{\partial u'}(u'_0, v'_0)\frac{du'}{dt}(t_0) + \frac{\partial v}{\partial v'}(u'_0, v'_0)\frac{dv'}{dt}(t_0).\end{aligned} \tag{11}$$

Comparison of (10) and (11) shows that this relation coincides with that for parameter transformation in the definition of a tangent plane.

The above definition of a tangent vector to a curve on a surface M is convenient in that the vector coordinates depend only on a local coordinate system (u, v) on M rather than on the way the surface M is embedded in a three-dimensional space $\mathbf{R}^3$. This statement can be formulated as the following lemma.

Lemma 1. *Let M be a non-degenerate surface in $\mathbf{R}^3$, $P_0 \in M$, (u, v) a local coordinate system in a neighbourhood of point P_0 on M, and $(u(t), v(t))$ a smooth curve on M. Then the tangent vector to the curve at P_0 has in the local coordinate system (u, v) the coordinates $\left(\frac{du}{dt}(t_0), \frac{dv}{dt}(t_0)\right)$. If $\mathbf{x} = \mathbf{x}(u, v)$ is parametric definition of M and $\mathbf{x}(t) = \mathbf{x}(u(t), v(t))$ is a curve, then*

$$\frac{d\mathbf{x}}{dt}(t_0) = \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\frac{\partial u}{\partial t}(t_0) + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\frac{dv}{dt}(t_0).$$

3.3.2. GENERAL DEFINITION OF A TANGENT VECTOR

The examples considered above show that it is more convenient to study the infinitesimal properties of a curve on a manifold using a local coordinate system. In particular, very important concepts in geometry are a tangent vector and a tangent space of an arbitrary smooth manifold, in complete analogy with the corresponding concepts for a surface in an ordinary three-dimensional space $\mathbf{R}^3$.

Definition 1. Let M be a smooth n-dimensional manifold and $P_0 \in M$ an arbitrary point. A *tangent vector* ξ *at point* P_0 *to the manifold* M is a correspondence which associates with any local coordinate system $(x_i^1, \ldots, x_i^n)$ a set of numbers $(\xi_i^1, \ldots, \xi_i^n)$ satisfying the following relation for each pair of local coordinate systems:

$$\xi_i^k = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l}(P_0)\, \xi_j^l. \tag{12}$$

The numbers $(\xi_i^1, \ldots, \xi_i^n)$ are called *coordinates of the tangent vector* ξ in the local coordinate system $(x_i^1, \ldots, x_i^n)$ and relation (12) is called a *tensor law of coordinate transformation* for the tangent vector ξ.

Definition 1 is a generalization of the concept of the coordinates of a tangent vector to a curve on a surface. The law (11) of the transformation of these coordinates is a particular case of the tensor law (12) of the coordinate transformation for a vector tangent to a manifold. Moreover, any smooth curve on a smooth manifold is endowed at each point with a tangent vector in the sense of Definition 1. This important property can be formulated as the following proposition.

Proposition 1. *Let M be a smooth manifold and $\gamma\colon (-1, 1) \to M$ a smooth mapping of the interval $(-1, 1)$ into M. Then the correspondence which associates with each local coordinate system $(x^1, \ldots, x^n)$ in a neighbourhood of point $P_0 = \gamma(0)$ the set of numbers $\left(\frac{dx^1}{dt}(\gamma(t)), \ldots, \frac{dx^n}{dt}(\gamma(t))\right)$, is a tangent vector in the sense of Definition 1.*

To *prove* Proposition 1, it is sufficient to verify the tensor law of coordinate transformation (12). Assume $\xi_j^k = \frac{dx_j^k}{dt}(\gamma(t))\Big|_{t=0}$, where $(x_j^1, \ldots, x_j^n)$ is a local coordinate system on M in a neighbourhood of point P_0. Then we obtain for two local coordinate systems

$$\xi_i^k = \frac{d}{dt} x_i^k(\gamma(t))|_{t=0} = \frac{d}{dt} x_i^k(x_j^1(\gamma(t)), \ldots, x_j^n(\gamma(t)))|_{t=0}$$

$$= \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l}(\gamma(t)) \frac{d}{dt} x_j^l(\gamma(t))|_{t=0} = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l}(P_0)\, \xi_j^l.$$

This is the tensor law (12). The proposition is proved.

It is therefore natural to call the correspondence used in Proposition 1 a *tangent vector to a curve* γ or a *velocity vector of a curve* γ. The tangent vector to a curve γ will be denoted by $\frac{d\gamma}{dt}(t_0)$ or $\dot{\gamma}(t_0)$.

3. 3. 3. TANGENT SPACE $T_{P_0}(M)$

The set of all tangent vectors to a manifold M at a fixed point P_0 is called a *tangent space to the manifold M at point P_0*. This set is denoted by $T_{P_0}(M)$. Each tangent vector $\xi \in T_{P_0}(M)$ is uniquely defined by its coordinates in a fixed coordinate system. Indeed, suppose we are given a set of numbers $(\eta^1, \ldots, \eta^n)$ and assume this set to be the coordinates of a tangent vector in question in a fixed local coordinate system $(x_{i_0}^1, \ldots, x_{i_0}^n)$, i.e. $\eta^k = \xi_{i_0}^k$.

In order to define a tangent vector, we have to find its coordinates in any local coordinate system $(x_i^1, \ldots, x_i^n)$. Let us put

$$\xi_i^k = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_{i_0}^l}(P_0)\,\eta^l.$$

These coordinates must satisfy the tensor law of coordinate transformation (12). To verify this, we substitute ξ_i^k and ξ_j^l into formula (12)

$$\sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_{i_0}^l}(P_0)\,\eta^l = \sum_{s=1}^{n} \frac{\partial x_i^k}{\partial x_j^s}(P_0) \sum_{l=1}^{n} \frac{\partial x_j^s}{\partial x_{i_0}^l}(P_0)\,\eta^l$$
$$= \sum_{l=1}^{n} \left(\sum_{s=1}^{n} \frac{\partial x_i^k}{\partial x_j^s}(P_0) \frac{\partial x_j^s}{\partial x_{i_0}^l}(P_0)\right)\eta^l.$$

Since $\frac{\partial x_i^k}{\partial x_{i_0}^l}(P_0) = \sum_{s=1}^{n} \frac{\partial x_i^k}{\partial x_j^s}(P_0)\, \frac{\partial x_j^s}{\partial x_{i_0}^l}(P_0)$ (the law of transformation of the Jacobi matrix under triple coordinate transformation), relation (12) is satisfied identically.

We have demonstrated that the set of all tangent vectors to a manifold M at point P_0 is uniquely determined by the coordinates of these vectors in one fixed local coordinate system. Hence, the entire tangent space $T_{P_0}(M)$ can be identified with the arithmetic vector space $\mathbf{R}^n$. This means that $T_{P_0}(M)$ can be endowed with the structure of a linear space. Seemingly, the structure of a linear space in $T_{P_0}(M)$ should depend on the local coordinate system in a neighbourhood of point P_0. This is not the case, however.

Proposition 2. *Addition of vectors and multiplication of a vector by a number in a tangent space $T_{P_0}(M)$ do not depend on the local coordinate system on M in a neighbourhood of point P_0.*

Proof. Let ξ, η be two vectors in $T_{P_0}(M)$ and $(x_i^1, \ldots, x_i^n)$, $(x_j^1, \ldots, x_j^n)$ two local coordinate systems on M in a neighbourhood of P_0. Suppose $(\xi_i^1, \ldots, \xi_i^n)$, $(\eta_i^1, \ldots, \eta_i^n)$ are the coordinates of the vectors ξ, η in the system $(x_i^1, \ldots, x_i^n)$ and $(\xi_j^1, \ldots, \xi_j^n)$, $(\eta_j^1, \ldots, \eta_j^n)$ are the coordinates of the same vectors in the system $(x_j^1, \ldots, x_j^n)$. By the tensor law (12) of coordinate transformation for tangent vectors we have the relations

$$\xi_i^k = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l}(P_0)\,\xi_j^l, \tag{13}$$

$$\eta_i^k = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l}(P_0)\,\eta_j^l. \tag{14}$$

Termwise addition of these relations yields

$$(\xi_i^k + \eta_i^k) = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l} (P_0) (\xi_j^l + \eta_j^l),$$

which means that the set $((\xi_i^1 + \eta_i^1), \ldots, (\xi_i^n + \eta_i^n))$ obey the tensor law of coordinate transformation, i.e. define the same tangent vector without any reference to a local coordinate system.

Similarly, multiplying (13) by a number λ, we obtain $\lambda\xi_i^k = \sum_{l=1}^{n} \frac{\partial x_i^k}{\partial x_j^l} (P_0) (\lambda\xi_j^l)$, i.e. the set of number $(\lambda\xi_i^k)$ also obeys the tensor law (12). The proposition is proved.

This tensor law of coordinate transformation may be considered as a method of identifying arithmetic spaces of the coordinates of tangent vectors in any local coordinate system. The method lies in multiplying the coordinate column (ξ_j^k) by the Jacobi matrix of the transformation from $(x_j^1, \ldots, x_j^n)$ to $(x_i^1, \ldots, x_i^n)$

$$\begin{pmatrix} \xi_i^1 \\ \cdot \\ \cdot \\ \cdot \\ \xi_i^n \end{pmatrix} = \begin{pmatrix} \frac{\partial x_i^1}{\partial x_j^1} & \cdots & \frac{\partial x_i^1}{\partial x_j^n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial x_i^n}{\partial x_j^1} & \cdots & \frac{\partial x_i^n}{\partial x_j^n} \end{pmatrix} \begin{pmatrix} \xi_j^1 \\ \cdot \\ \cdot \\ \cdot \\ \xi_j^n \end{pmatrix}$$

or $(\xi_i^k) = \left(\frac{\partial x_i}{\partial x_j}\right) (\xi_j^l)$.

Hence, the tangent space $T_{P_0}(M)$ is a space isomorphic to all arithmetic spaces of the coordinates of tangent vectors.

3.3.4. SHEAF OF TANGENT CURVES

In previous subsections we have introduced the formal algebraic definition of a tangent space to a manifold. This definition, however, does not reflect an obvious geometric property of a tangent vector—linear approach to a curve. How can we define this property without any reference to the position of a manifold in a linear space? A vector may correspond to many curves which it touches. Thus, we have nothing to do but choose such classes of curves which have a common tangent vector in linear spaces.

Definition 2. Two curves γ_1 and γ_2 on a manifold M intersecting at a point P_0 are called *tangent* if in each local coordinate system

$(x^1, \ldots, x^n)$ in a neighbourhood of P_0 we have

$$\sum_{k=1}^{n} (x^k(\gamma_1(t)) - x^k(\gamma_2(t)))^2 = o(t-t_0)^2 \text{ for } t \to t_0. \tag{15}$$

As before, it is sufficient to verify the tangency condition (15) only in one local coordinate system. The tangency condition is closely related to tangent vectors. In particular the following theorem justifies the term "tangent curves".

Theorem 1. *Two smooth curves γ_1 and γ_2 on a manifold M are tangent at a point P_0 if and only if their tangent vectors at P_0 coincide.*

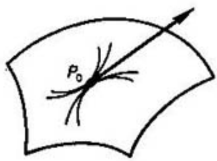

Figure 3.10

Proof. Condition (15) can be rewritten as

$$\lim_{t\to t_0} \sum_{k=1}^{n} \left(\frac{x^k(\gamma_1(t)) - x^k(\gamma_2(t))}{t-t_0}\right)^2 = 0.$$

After elementary manipulations we obtain

$$\sum_{k=1}^{n} \left(\frac{d}{dt} x^k(\gamma_1(t)) - \frac{d}{dt} x^k(\gamma_2(t))\right)^2 \Big|_{t=t_0} = 0$$

or

$$\frac{d}{dt} x^k(\gamma_0(t))|_{t=t_0} = \frac{d}{dt} x^k(\gamma_2(t))|_{t=t_0}.$$

The last equality exactly means that the tangent vectors to the curves γ_1 and γ_2 coincide, $\dot{\gamma}_1(t_0) = \dot{\gamma}_2(t_0)$.

Conversely, if $\dot{\gamma}_1(t_0) = \dot{\gamma}_2(t_0)$, we have

$$\lim_{t\to t_0} \sum_{k=1}^{n} \left(\frac{x^k(\gamma_1(t)) - x^k(\gamma_2(t))}{t-t_0}\right)^2$$
$$= \sum_{k=1}^{n} \left(\frac{d}{dt} x^k(\gamma_1(t)) - \frac{d}{dt} x^k(\gamma_2(t))\right)^2 = 0.$$

Theorem 1 gives an alternative definition of a tangent vector to a curve. The set of all smooth curves through a given point P_0 on a manifold M splits into disjoint classes of pairwise tangent curves. Let the class of tangent curves through a point $P_0 \in M$ be called a *tangent vector*. We have then a one-to-one correspondence between tangent vectors in the sense of Definition 1 and in the sense of classes of tangent curves (Fig. 3.10).

3.3.5. DIRECTIONAL DERIVATIVE OF A FUNCTION

There is one more way of representing a tangent vector on a manifold M. We shall start, as usual, with a simple example. Let $f(x, y)$ be a smooth function of two variables, $P_0 = (x_0, y_0)$

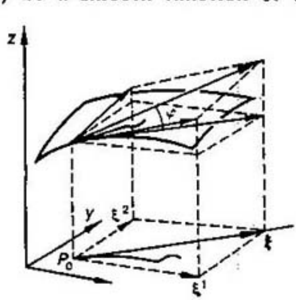

Figure 3.11

a point, and $\xi = (\xi^1, \xi^2)$ a vector in a plane $\mathbf{R}^2$. In mathematical analysis we often deal with the derivative of a function f with respect to a vector ξ, which is defined as

$$\xi(f) = \frac{\partial f}{\partial x}(x_0, y_0)\,\xi^1 + \frac{\partial f}{\partial y}(x_0, y_0)\,\xi^2. \tag{16}$$

The derivative with respect to the vector ξ can also be defined in terms of a smooth curve. Let $\gamma(t)$ be a smooth curve through a point P_0 on $\mathbf{R}^2$. Suppose the tangent vector to $\gamma(t)$ at P_0 is ξ. We have then (see Fig. 3.11)

$$\xi(f) = \frac{d}{dt} f(\gamma(t))\big|_{t=t_0}, \quad \gamma(t_0) = P_0. \tag{17}$$

Indeed, if $\gamma(t) = (x(t), y(t))$, then $\xi^1 = \frac{dx}{dt}(t_0)$, $\xi^2 = \frac{dy}{dt}(t_0)$. Substituting the coordinates of the curve $\gamma(t)$ into the arguments of f

and differentiating with respect to the parameter, we obtain

$$\begin{aligned}\frac{d}{dt} f(\gamma(t))|_{t=t_0} &= \frac{d}{dt} f(x(t), y(t))|_{t=t_0}\\ &= \frac{\partial f}{\partial x}(x(t_0), y(t_0))\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(x(t_0), y(t_0))\frac{dy}{dt}(t_0)\\ &= \frac{\partial f}{\partial x}(x_0, y_0)\xi^1 + \frac{\partial f}{\partial y}(x_0, y_0)\xi^2 = \xi(f).\end{aligned}$$

We now define differentiation of a smooth function at a point P_0 with respect to each tangent vector ξ to a manifold M at $P_0 \in M$.

Definition 3. Suppose $P_0 \in M$, $\xi \in T_{P_0}(M)$, $\gamma(t)$ is a smooth curve through P_0, $\gamma(t_0) = P_0$, ξ is its tangent vector at P_0, $\dot{\gamma}(t_0) = \xi$, and f is a smooth function on M. The derivative

$$\frac{d}{dt} f(\gamma(t))|_{t=t_0} = \xi(f) \tag{18}$$

is called the *derivative of the function f with respect to the tangent vector* ξ. Calculation of the derivative is called the *differentiation of the function with respect to the vector* ξ.

Theorem 2. *Let $(x^1, \ldots, x^n)$ be a local coordinate system in a neighbourhood of a point $P_0 = (x_0^1, \ldots, x_0^n)$ of a manifold M, let $\xi = (\xi^1, \ldots, \xi^n)$ be a tangent vector to M at P_0, and let $f = f(x^1, \ldots, x^n)$ be a smooth function in the neighbourhood of P_0 represented as a function of the local coordinates $(x^1, \ldots, x^n)$. Then*

$$\xi(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\xi^i. \tag{19}$$

Hence, the definition (18) *of the derivative does not depend on the choice of a curve in the class of tangent curves, and the right-hand side of* (19) *does not depend on a local coordinate system. If g is another smooth function in the neighbourhood of P_0, the product of f and g obeys the Leibniz formula*

$$\xi(fg) = f(x_0^1, \ldots, x_0^n)\xi(g) + \xi(f) g(x_0^1, \ldots, x_0^n). \tag{20}$$

Proof. We represent the curve $\gamma(t)$ through P_0 in the coordinate form: $\gamma(t) = (x^1(t), \ldots, x^n(t))$. By the definition of a tangent vector to a curve, we have $\frac{dx^i}{dt}(t_0) = \xi^i$, whence

$$\begin{aligned}\xi(f) &= \frac{d}{dt} f(\gamma(t))|_{t=t_0} = \frac{d}{dt} f(x^1(t), \ldots, x^n(t))|_{t=t_0}\\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n) \left.\frac{dx^i}{dt}\right|_{t=t_0} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\xi^i.\end{aligned}$$

Formula (19) is proved. To prove formula (20), we use (19) and differentiate the product of functions to obtain

$$\begin{aligned}
\xi(fg) &= \sum_{i=1}^{n} \frac{\partial}{\partial x^i}\left(f(x^1, \ldots, x^n)\, g(x^1, \ldots, x^n)\right)\Big|_{x^k=x_0^k} \xi^i \\
&= \sum_{i=1}^{n} \Big(\frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, g(x_0^1, \ldots, x_0^n) \\
&\qquad + f(x_0^1, \ldots, x_0^n)\, \frac{\partial g}{\partial x^i}(x_0^1, \ldots, x_0^n) \Big) \xi^i \\
&= \Big(\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, \xi^i \Big)\, g(x_0^1, \ldots, x_0^n) \\
&\qquad + f(x_0^1, \ldots, x_0^n) \Big(\sum_{i=1}^{n} \frac{\partial g}{\partial x^i}(x_0^1, \ldots, x_0^n)\, \xi^i \Big) \\
&= \xi(f)\, g(x_0^1, \ldots, x_0^n) + f(x_0^1, \ldots, x_0^n)\, \xi(g).
\end{aligned}$$

Thus, formula (20) is also proved.

Differentiation of a smooth function f with respect to a tangent vector ξ is characterized by certain properties which do not involve any local coordinate system. This way of presentation is most convenient to demonstrate the independence of our geometric constructions of a particular local coordinate system. There are two such properties:

(a) differentiation with respect to a vector ξ is linear, i.e. if f and g are smooth functions, and λ and μ are arbitrary numbers, then

$$\xi(\lambda f + \mu g) = \lambda \xi(f) + \mu \xi(g), \tag{21}$$

(b) differentiation with respect to a vector ξ satisfies the Leibniz formula (20).

We may give a general definition.

Definition 4. The operation A which associates with any smooth function f of class C^∞ on a smooth manifold M the number $A(f)$ satisfying (a) and (b) is called *differentiation at points* $P_0 \in M$.

In the Leibniz formula (20) the values of functions are calculated at a single point P_0, so that differentiation at distinct points P_0 and P_1 need not coincide. Obviously, differentiation with respect to a tangent vector ξ is a particular case of differentiation in the sense of Definition 4, and there are no other differentiation operations. This means that for each differentiation in the sense of Definition 4 there exists a tangent vector with respect to which the function is differentiated.

Theorem 3. *Let M be a C^∞-manifold, $P_0 \in M$ be an arbitrary point, and let A denote differentiation in the sense of Definition 4. Then there exists a unique tangent vector ξ at P_0 such that $A(f) = \xi(f)$ for any smooth function f in a neighbourhood of P_0.*

Proof. The tangent vector will be sought as a column of its coordinates in some local coordinate system $(x^1, \ldots, x^n)$ in a neighbourhood of P_0. Then any smooth function will be represented as a function of the variables $(x^1, \ldots, x^n)$.

Lemma 2. *Any C^∞-function $f(x^1, \ldots, x^n)$ can be represented in the form*

$$f(x^1, \ldots, x^n) = f(x_0^1, \ldots, x_0^n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)(x^i - x_0^i)$$

$$+ \sum_{i,j=1}^{n} h_{ij}(x^1, \ldots, x^n)(x^i - x_0^i)(x^j - x_0^j), \qquad (22)$$

where $h_{ij}(x^1, \ldots, x^n)$ are C^∞-functions.

Proof. Write the identity

$$f(x^1, \ldots, x^n) \equiv f(x_0^1, \ldots, x_0^n) + \int_0^1 \frac{d}{dt} f(x_0^1 + t(x^1 - x_0^1), \ldots, x_0^n + t(x^n - x_0^n))\, dt$$

and differentiate it with respect to t under the integral sign

$$f(x^1, \ldots, x^n) = f(x_0^1, \ldots, x_0^n)$$

$$+ \int_0^1 \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1 + t(x^1 - x_0^1), \ldots, x_0^n + t(x^n - x_0^n))(x^i - x_0^i)\, dt$$

$$= f(x_0^1, \ldots, x_0^n) + \sum_{i=1}^{n} (x^i - x_0^i) h_i(x^1, \ldots, x^n). \qquad (23)$$

In the last equality the functions

$$h_i(x^1, \ldots, x^n) = \int_0^1 \frac{\partial f}{\partial x^i}(x_0^1 + t(x^1 - x_0^1), \ldots, x_0^n + t(x^n - x_0^n))\, dt$$

are smooth of class C^∞. Substitution of $x^i = x_0^i$ yields

$$h_i(x_0^1, \ldots, x_0^n) = \int_0^1 \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, dt = \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n). \qquad (24)$$

Applying formula (23) to the functions $h_i(x^1, \ldots, x^n)$, we obtain

$$h_i(x^1, \ldots, x^n) = h_i(x_0^1, \ldots, x_0^n) + \sum_{j=1}^{n} (x^j - x_0^j)\, h_{ij}(x^1, \ldots, x^n), \tag{25}$$

where $h_{ij}(x^1, \ldots, x^n)$ are C^∞-functions. Substituting (25) into (23) and taking into account (24), we arrive at the initial representation (22).

Lemma 3. *Let f and g be smooth functions on a manifold M such that $f(P_0) = g(P_0) = 0$. Then for any differentiation A at P_0 the equality $A(fg) = 0$ is satisfied.*

The *proof* of Lemma 3 directly follows from the Leibniz formula (20).

Let us now turn to the proof of Theorem 3. Represent f in the form of (22) and apply differentiation A to the left-hand and right-hand sides. Since the operation is linear, we have

$$A(f) = f(x_0^1, \ldots, x_0^n)\, A(1) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, A(x^i - x_0^i)$$
$$+ \sum_{i,j=1}^{n} A((x^i - x_0^i)(x^j - x_0^j)\, h_{ij}(x^1, \ldots, x^n)). \tag{26}$$

Note that for a constant unit function we have $A(1) = A(1\cdot 1) = A(1)\cdot 1 + 1\cdot A(1) = 2A(1) = 0$. In the last sum of (26) each term can be represented as the product of two functions $(x^i - x_0^i)$ and $(x^j - x_0^j)\, h_{ij}(x^1, \ldots, x^n)$, each vanishing at $P_0 \in M$. Thus, by Lemma 3, $A((x^i - x_0^i)(x^j - x_0^j)\, h_{ij}(x^1, \ldots, x^n)) = 0$. Hence,

$$A(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, A(x^i - x_0^i). \tag{27}$$

Since in formula (27) the function f is arbitrary, by putting $\xi^i = A(x^i - x_0^i)$, we obtain the vector $\xi = (\xi^1, \ldots, \xi^n)$ such that

$$A(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, \xi^i = \xi(f).$$

We now prove that the vector ξ is unique. If we could find two distinct vectors ξ and η such that $A(f) = \xi(f) = \eta(f)$, then assuming $\zeta = \xi - \eta \neq 0$ we would obtain $\zeta(f) = 0$ for any C^∞-function f in a neighbourhood of $P_0 \in M$. This cannot be the case, however. Indeed, in the local coordinate system $(x^1, \ldots, x^n)$ the vector ζ has the coordinates $(\zeta^1, \ldots, \zeta^n)$ not all of which are zero. Let

$\zeta^k \neq 0$. Then we obtain for the function $f(x^1, \ldots, x^n) \equiv x^k$

$$\zeta(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\, \zeta^k = \sum_{i=1}^{n} \frac{\partial x^k}{\partial x^i}\, \zeta^k = \zeta^k \neq 0.$$

This completes the proof of Theorem 3.

The theorem establishes a one-to-one correspondence between tangent vectors to a manifold M at a point $P_0 \in M$ and differentiation of a smooth function at P_0. We can therefore formulate a third equivalent definition of a tangent vector: a *tangent vector* is a differential operator applied to a smooth function at point P_0 of a manifold M.

Calculation of a partial derivative in a local coordinate system $(x^1, \ldots, x^n)$ may serve as an example of differentiation of a smooth function. By Theorem 3, the operator $\frac{\partial}{\partial x^k}$ is a tangent vector with the coordinates $(0, \ldots, 1, \ldots, 0)$ where 1 is the kth coordinate. Thus, the tangent vectors $\left\{\frac{\partial}{\partial x^k}\right\}$ form a basis in the tangent space $T_P(M)$, and any tangent vector $\xi = (\xi^1, \ldots, \xi^n)$ can be decomposed into the linear combination $\xi = \xi^1 \frac{\partial}{\partial x^1} + \ldots + \xi^n \frac{\partial}{\partial x^n}$. This convenient representation will often be used below.

3.3.6. TANGENT BUNDLES

We have already seen that the set of all tangent vectors $T_{P_0}(M)$ to a manifold M at a point P_0 is a linear space of the same dimension as that of M. In geometry it is sometimes useful to study the whole set of tangent vectors to a manifold M, which can, apparently, be represented as the union $\bigcup_{P_0 \in M} T_{P_0}(M)$. This space (not yet topological) is denoted by $T(M)$ and called a tangent bundle of M. The term bundle means that $T(M)$ consists of fibres—tangent spaces $T_{P_0}(M)$ to distinct points P_0 of the manifold M. A tangent bundle is by no means a vector space, for it is meaningless to add vectors belonging to different fibres. If, for example, a manifold M is a two-dimensional surface in $\mathbf{R}^3$, then $T(M)$ represents the union of all tangent planes to M. It should be noted that tangent planes to a surface usually intersect, that is, they have common points. According to the definition of $T(M)$, however, these points in each fibre define different vectors, since they originate at distinct points.

Let us consider a circle $S^1 \subset \mathbf{R}^2$. Figure 3.12 depicts two tangents to the circle at points P_0 and Q_0 and two tangent vectors ξ and η with common end. Therefore, in order to describe the tangent bundle $T(S^1)$ as a topological space in a Euclidean space, we have to go over to a three-dimensional space $\mathbf{R}^3$ and "rotate" the tangents through

an angle relative to the plane (x, y) in such a way that they do not intersect (see Fig. 3.13). The tangent bundle $T(S^1)$ becomes then a one-sheet hyperboloid (geometric cylinder). Under such an operation the property of "close approach" of $T_{P_0}(S^1)$ to the circle is

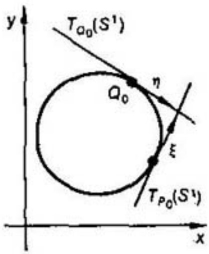

Figure 3.12

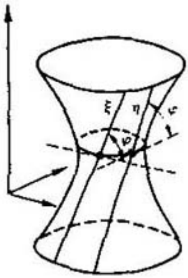

Figure 3.13

lost. The situation is as if we have "torn off" the tangent to the circle and have disregarded that the fibres must be tangent to S^1. We may slightly modify the embedding of the tangent bundle in $\mathbf{R}^3$, so that $T_{P_0}(S^1)$ remains tangent to S^1 and different fibres do not inter-

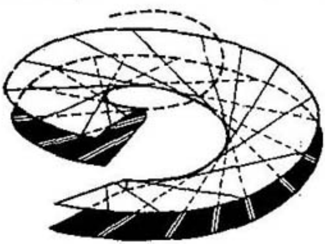

Figure 3.14

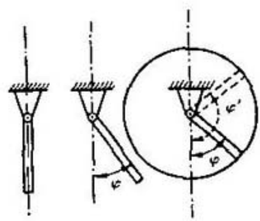

Figure 3.15

sect. This cannot be done on the entire circle S^1, but only on its part, an arc. Let us embed an arc in $\mathbf{R}^3$ as a helix. Then tangents to the helix do not intersect (Fig. 3.14).

Here are some examples from mechanics which demonstrate that non-trivial manifolds and tangent bundles to them are convenient in the description of mechanical systems.

Example 1. Consider the motion of a plane pendulum, i.e. a rigid bar hinged at a point (Fig. 3.15). The position of the bar is deter-

mined by one parameter, the angle φ between the bar axis and the vertical. Thus, the set of all positions of the bar is a circle S^1. Such a set is called a *configurational space*.

Consider a two-link compound pendulum, i.e. two bars pivoted together (Fig. 3.16). The position of this pendulum is determined by two angles φ_1 and φ_2, so that the set of all positions represents a two-dimensional torus $T^2 = S^1 \times S^1$. Figure 3.17 depicts another system whose configurational space is a torus embedded in $\mathbf{R}^3$.

Example 2. In mechanics the motion of a mechanical system is usually described by the parameters characterizing the position of

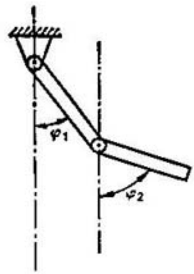

Figure 3.16

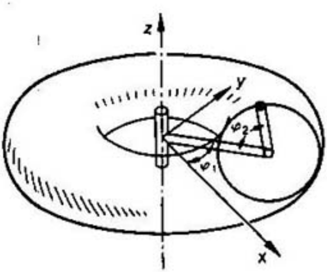

Figure 3.17

the system and by the velocities of its parts. The set of all these positions and velocities is called a *phase space*, which can naturally be identified with a tangent bundle to a configurational space. For instance, if a particle moves along a two-dimensional sphere at a constant velocity, the phase space is a subset in a tangent bundle consisting of tangent vectors of constant length.

Example 3. There exist more complicated configurational and phase spaces. Let us consider, for instance, a three-dimensional rigid body with a fixed point. Any position of this body in $\mathbf{R}^3$ can be described as follows. Choose in the body three orthogonal unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ emerging from the fixed point. Any position of the body is given then by the position of these three vectors in $\mathbf{R}^3$. Thus, the configurational space can be identified with a connectedness component of the set of all orthogonal unit bases in $\mathbf{R}^3$.

3.4. SUBMANIFOLDS

We now turn to differential calculus on a smooth manifold. Many important concepts of mathematical analysis, such as the differential of a function, critical points, implicit functions, can

naturally be extended to an arbitrary smooth manifold. Within the framework of a general theory of smooth manifolds we can easily interpret many concepts of differential calculus: the differential of a function, the gradient of a function, the implicit function theorem, regular points of a function, etc.

3.4.1. DIFFERENTIAL OF A SMOOTH MANIFOLD

The concept of the differential of a smooth manifold can readily be transferred to an arbitrary smooth mapping of manifolds.

Definition 1. Let $f: M_1 \to M_2$ be a smooth mapping of smooth manifolds. The *differential* df_{P_0} *of a smooth mapping* f *at a point* $P_0 \in M_1$ is a linear mapping of a tangent space $T_{P_0}(M_1)$ into a tangent space $T_{Q_0}(M_2)$, $Q_0 = f(P_0)$, defined in local coordinate systems by the Jacobi matrix of f.

Recall that in Sec. 3.1 we defined the Jacobi matrix for a system of functions $y^1 = f^1(x^1, \ldots, x^n)$, $\ldots$, $y^m = f^m(x^1, \ldots, x^n)$ as a matrix of partial derivatives

$$df = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x^n} \\ \vdots & & \\ \frac{\partial f^m}{\partial x_1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{pmatrix}.$$

Then, if $(x^1, \ldots, x^n)$ is a local coordinate system on M_1 in a neighbourhood of point P_0 and $(y^1, \ldots, y^m)$ is a local coordinate system on M_2 in a neighbourhood of Q_0, the mapping f can be represented as a set of coordinate functions $y^k = f^k(x^1, \ldots, x^n)$, and $T_{P_0}(M_1)$ and $T_{Q_0}(M_2)$ can be represented as arithmetic spaces of columns of length n and m, respectively. Let the tangent vector $\xi \in T_{P_0}(M_1)$ have coordinates $(\xi^1, \ldots, \xi^n)$ and the vector $\eta \in T_{Q_0}(M_2)$ coordinates $(\eta^1, \ldots, \eta^m)$. If $\eta = df_{P_0}(\xi)$, then

$$\begin{pmatrix} \eta^1 \\ \vdots \\ \eta^m \end{pmatrix} = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(P_0) \cdots \frac{\partial f^1}{\partial x^n}(P_0) \\ \vdots \\ \frac{\partial f^m}{\partial x^1}(P_0) \cdots \frac{\partial f^m}{\partial x^n}(P_0) \end{pmatrix} \begin{pmatrix} \xi^1 \\ \vdots \\ \xi^n \end{pmatrix}. \qquad (1)$$

Seemingly, Definition (1) should depend on the choice of a local coordinate system in the neighbourhoods of $P_0 \in M_1$ and $Q_0 \in M_2$. This is not the case, however. Indeed, we can choose another pair of local coordinate systems in the neighbourhoods of $P_0 \in M_1$ and $Q_0 \in M_2$ and verify, using the tensor law of coordinate transformation for tangent vectors, that Definition 1 is invariant. Yet, we shall proceed differently. We have introduced three definitions of a tangent vector to a manifold M. One of the definitions, namely that

employing differentiation of a function, does not involve local coordinates. Thus, if we reformulate Definition 1 in terms of differentiation of a smooth function, we automatically arrive at the independence of the differential of the mapping f of local coordinate systems in manifolds M_1 and M_2.

Lemma 1. *Let $f\colon M_1 \to M_2$ be a smooth mapping, $f(P_0) = Q_0$, $\xi \in T_{P_0}(M_1)$ be a tangent vector to M_1 at point P_0, and $\eta = df_{P_0}(\xi)$ be a tangent vector to M_2 at point Q_0 in the sense of Definition 1. Then for any smooth function g on M_2 the following relation is satisfied:*

$$\eta(g) = \xi(g \circ f). \tag{2}$$

Proof. According to Definition 1, we have to choose local coordinate systems $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^m)$ in M_1 and M_2 in the neighbourhoods of points P_0 and $Q_0 = f(P_0)$, respectively. In this case f is represented as a set of functions $y^k = f^k(x^1, \ldots, x^n)$, the function g is replaced by a smooth function $g = g(y^1, \ldots, y^m)$, the vectors ξ and η acquire the coordinates $(\xi^1, \ldots, \xi^n)$ and $(\eta^1, \ldots, \eta^m)$, respectively. We have then

$$\eta(g) = \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0)\, \eta^l,$$

$$\begin{aligned}
\xi(g\circ f) &= \sum_{j=1}^{n} \frac{\partial}{\partial x^j}(g\circ f)\, \xi^j \\
&= \sum_{j=1}^{n} \frac{\partial}{\partial x^j}\left(g\left(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)\right)\right) \xi^j \\
&= \sum_{j=1}^{n} \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0)\, \frac{\partial f^l}{\partial x^j}(P_0)\, \xi^j \\
&= \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0) \left(\sum_{j=1}^{n} \frac{\partial f^l}{\partial x^j}(P_0)\, \xi^j \right).
\end{aligned}$$

Applying formula (1) to the last equality, we obtain

$$\xi(g\circ f) = \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0)\, \eta^l = \eta(g).$$

Lemma 1 is proved.

Considering relation (2) as the definition of the differential of a mapping f, i.e. assuming

$$df_{P_0}(\xi)(g) = \xi(g \circ f), \tag{3}$$

we can infer from Lemma 1 that the operation $g \to df_{P_0}(\xi)(g)$ is differentiation of a function on M_2 at p⌃'n' Q_0, i.e. $df_{P_0}(\xi)$ is a tangent vector to M_2 at point Q_0; this conclusion coincides with Definition 1 by formula (1). Hence, Definition 1 does not depend on the choice of a local coordinate system.

Let us finally consider the last definition of a tangent vector in terms of a sheaf of tangent curves.

Lemma 2. *Let $f\colon M_1 \to M_2$ be a smooth mapping, $Q_0 = f(P_0)$, γ be a smooth curve in M_1 through P_0, and g be a smooth function on M_2. Let also $\xi = \dot{\gamma}(t_0)$ and $\eta = df_{P_0}(\xi)$. Then*

$$\eta(g) = \frac{d}{dt} g(f(\gamma(t)))|_{t=t_0}. \tag{4}$$

It is most convenient to *prove* Lemma 2 in the local coordinate systems $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^m)$ of manifolds M_1 and M_2, respectively. Suppose that all the functions and vectors have the same coordinate representations as in Lemma 1, and $\gamma(t) = (x^1(t), \ldots, x^n(t))$, $\xi^k = \frac{dx^k}{dt}(t_0)$. Then,

$$\eta(g) = \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0)\, \eta^l,$$

$$\frac{d}{dt} g(f(\gamma(t)))|_{t=t_0} = \sum_{l,j} \frac{\partial g}{\partial y^l}(Q_0) \frac{\partial f^l}{\partial x^j}(P_0) \frac{dx^j}{dt}(t_0)$$

$$= \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0) \sum_{j=1}^{n} \frac{\partial f^l}{\partial x^j}(P_0)\, \xi^j = \sum_{l=1}^{m} \frac{\partial g}{\partial y^l}(Q_0)\, \eta^l = \eta(g).$$

Lemma 2 implies that if a tangent vector ξ is represented as a tangent vector of a curve $\gamma(t)$, $\dot{\gamma}(t_0) = \xi \in T_{P_0}(M)$, the image $\eta = df_{P_0}(\xi)$ is represented as a tangent vector of the curve $f(\gamma(t))$ on manifold M_2. Furthermore, if follows from Lemma 2 that if the curves $\gamma_1(t)$ and $\gamma_2(t)$ are tangent at point $P_0 \in M_1$, their images $f(\gamma_1(t))$ and $f(\gamma_2(t))$ are tangent at point $Q_0 \in M_2$. The differential of the mapping f can therefore be defined as a mapping which to a sheaf of tangent curves $\{\gamma(t)\}$ at point P_0 associates a sheaf of tangent curves at point Q_0 of manifold M_2, the sheaf containing all the curves $\{f(\gamma(t))\}$. This definition is not so clear as to reveal why the differential df_{P_0} is a *linear mapping* of tangent spaces. Figure 3.18 shows the mapping $f(x, y) = (x, y^2)$, $f\colon \mathbf{R}^2 \to \mathbf{R}^2$, and it can be seen that the sheaf of tangent curves at point $P_0 = (0, 0)$ does not cover under the mapping f the entire sheaf of tangent curves at point Q_0.

Example 1. Let us consider a smooth function $y = f(x)$ as a smooth mapping of manifolds $f\colon \mathbf{R}^1 \to \mathbf{R}^1$. Then, according to Definition 1,

the differential is a linear mapping of $T_x(\mathbf{R}^1) = \mathbf{R}^1$ into $T_{f(x)}(\mathbf{R}^1) = \mathbf{R}^1$ defined by formula (1), i.e. $\eta = f'(x)\,\xi$. In mathematical analysis by the differential of a function $f(x)$ we mean the linear part of the increment in f as a function of two independent arguments: $dy = f(x)\,dx$. Thus, assuming $dx = \xi$ and $dy = \eta$, we find that the concepts coincide.

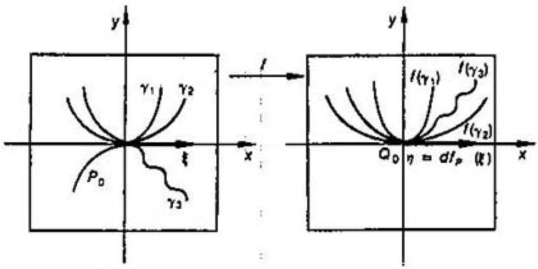

Figure 3.18

Example 2. Consider a smooth function f of n independent variables $y = f(x^1, \ldots, x^n)$. Just like in the case of one variable, we shall represent f as a smooth mapping of manifolds $f: \mathbf{R}^n \to \mathbf{R}^1$. Then the differential of the mapping f is a linear mapping of tangent

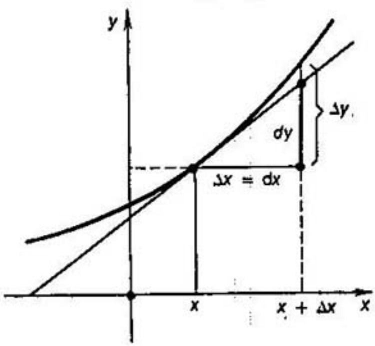

Figure 3.19

spaces $df_{P_0}: T_{P_0}(\mathbf{R}^n) = \mathbf{R}^n \to T_{Q_0}(\mathbf{R}^1) = \mathbf{R}^1$, where $P_0 = (x_0^1, \ldots, x_0^n)$, $Q_0 = f(x_0^1, \ldots, x_0^n)$. The differential df_{P_0} in local coordinates is given by $\eta = \sum_{k=1}^{n} \frac{\partial f}{\partial x^k}(P_0)\,\xi^k$. On the other hand, the

differential of a function f is the linear part of the increment in f as a function of two groups of independent variables, $(x^1, \ldots, x^n)$ and $(dx^1, \ldots, dx^n)$, i.e.

$$dy = df(x^1, \ldots, x^n) = \sum_{k=1}^{n} \frac{\partial f}{\partial x^k}(x^1, \ldots, x^n)\, dx^k.$$

Assuming $\xi^k = dx^k$ and $\eta = dy$, we again find that the concept of the differential of a function coincides with the concept of the differential of a function represented as a mapping of manifolds. Furthermore, the matrix of df_{P_0} is the Jacobi matrix of the mapping f and, on the other hand, it is the gradient of the function f, that is, grad f is the matrix of df_{P_0} in a fixed coordinate system. Obviously, if a coordinate system is modified, the components of grad f will also change.

Example 3. Let us consider a smooth function $f = f(x^1, \ldots, x^n)$ and let $P_0 = (x_0^1, \ldots, x_0^n)$ be an extremal point of f. Then, by a

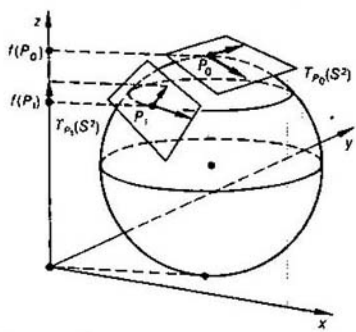

Figure 3.20

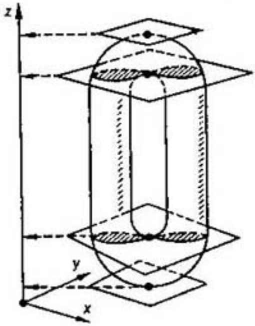

Figure 3.21

theorem of mathematical analysis, grad $f_{P_0} = 0$. In our terminology this means $df_{P_0} = 0$. This statement can be extended to an arbitrary manifold: *if a smooth function f has a local maximum at a point $P_0 \in M$, then $df_{P_0} = 0$.* This circumstance finds a new interpretation in the theory of manifolds. In the case of a Euclidean domain there always exists a smooth function f such that grad $f \neq 0$ at each point. This is not however the case for a smooth manifold. For example, for any smooth function f on a two-dimensional sphere S^2 the differential df vanishes at least at two points: at the maximum and at the minimum of f. In general, *if M is a compact smooth manifold, the differential df of a smooth function vanishes at least at two points.* Figure 3.20 depicts the behaviour of the differential of the "altitude"

function on a two-dimensional sphere S^2, and Fig. 3.21 shows the "altitude" function on a torus T^2.

Example 4. Let $f\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear mapping which is expressed in coordinates as the matrix $Y = AX$,

$$Y = \begin{pmatrix} y^1 \\ \cdot \\ \cdot \\ y^m \end{pmatrix}, \quad X = \begin{pmatrix} x^1 \\ \cdot \\ \cdot \\ x^n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

Then, obviously, the Jacobi matrix of the mapping f (and hence the differential df_{P_0}) coincides with A. In other words, $df_{P_0}(\xi) = f(\xi)$. And it is no surprise. If we consider the differential df_{P_0} as the linear part of the increment of the mapping, just like in the case of a function of many variables, and assume $\Delta X = \xi$, we obtain $\Delta Y = A(X + \Delta X) - AX = A\Delta X$. Hence, the increment ΔY coincides identically with its linear part, which means that the differential of a linear mapping does not depend on point $P_0 \in \mathbf{R}^n$.

3.4.2. DIFFERENTIAL AND LOCAL PROPERTIES OF MAPPINGS

From mathematical analysis we know the important property of a smooth function: using differential properties of a function at a point, we can deduce analogous properties in a neighbourhood of the point (though the neighbourhood can be small). For example, if the derivative $f'(x_0)$ of a function f is positive, $f'(x_0) > 0$, the function f increases in a small neighbourhood of x_0. A question naturally arises: what is the analogy between the positive sign of the derivative $f'(x_0)$ and an increase of f in a neighbourhood of point x_0? This analogy becomes even more obvious if the condition $f'(x_0) > 0$ is formulated as follows: the differential of a function f at a point x_0 is an increasing (linear) function. The corresponding theorem of mathematical analysis then reads: *if the differential of a function f at a point x_0 strictly increases, the function f itself also strictly increases in a neighbourhood of x_0*. As a second example, we may consider the implicit function theorem: the equation $f(x, y) = 0$ has a solution $y = y(x)$ if $\frac{\partial}{\partial y} f(x_0, y_0) \neq 0$ at a point (x_0, y_0) for which $f(x_0, y_0) = 0$. Let us now formulate this theorem in terms of differentials. The differential of a function f at a point (x_0, y_0) is

$$df = \frac{\partial f}{\partial x}(x_0, y_0)\,dx + \frac{\partial f}{\partial y}(x_0, y_0)\,dy.$$

Replacing in the equation $f(x, y) = 0$ the function f by its differential, we arrive at a new equation $df_{(x_0, y_0)} = 0$ which in the new vari-

ables (dx, dy) is linear and has the form

$$\frac{\partial f}{\partial x}(x_0, y_0)\,dx+\frac{\partial f}{\partial y}(x_0, y_0)\,dy=0. \tag{5}$$

The condition $\frac{\partial f}{\partial y}(x_0, y_0)\neq 0$ is then equivalent to the existence of a solution of Eq. (5). Thus, the implicit function theorem can be formulated as follows: *if the equation in terms of differentials* (5) *has a solution, the initial equation* $f(x, y) = 0$ *also has a solution in a neighbourhood of point* (x_0, y_0).

This leads us to a principle which proves to be useful in an analysis of smooth manifolds and their mappings: *if a property of the differential of a mapping of smooth manifolds is satisfied at some point, the analogous property holds true for the mapping itself in a neighbourhood of the same point.*

This principle is usually called the *principle of general position* or the *principle of linearization of a mapping*. The first term implies that the property of a linear mapping is preserved under small nonlinear distortions of the mapping. Naturally, this principle does not hold for each mapping, and we shall not attempt at elucidating the exact region of its applicability. We shall concentrate here on the most important theorems relating the properties of a mapping and its differential.

We shall start with the implicit function theorem for manifolds.

Definition 2. Let $f\colon M_1 \to M_2$ be a smooth mapping. The point $P_0 \in M_1$ is called a *regular point of* f if the differential of the mapping $df_{P_0}\colon T_{P_0}(M_1) \to T_{Q_0}(M_2)$, $Q_0 = f(P_0)$, is an epimorphism, i.e. a mapping onto the entire space $T_{Q_0}(M_2)$. The point $Q_0 \in M_2$ is called a *regular point* of the mapping f if any point P_0 of the inverse image $f^{-1}(Q_0)$ is a regular point of f.

This definition is, in fact, the implicit function theorem formulated in terms of the differential of a mapping. Indeed, in the local coordinate systems $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^m)$ in the neighbourhoods of points P_0 and Q_0, respectively, the mapping f is written as the following system of functions $y^k = f^k(x^1, \ldots, x^n)$:

$$\begin{aligned} f^1(x^1, \ldots, x^n) &= y_0^1 \\ &\;\;\vdots \\ f^m(x^1, \ldots, x^n) &= y_0^m. \end{aligned} \tag{6}$$

Since $f(P_0) = Q_0$, $P_0 = (x_0^1, \ldots, x_0^n)$, $Q_0 = (y_0^1, \ldots, y_0^m)$, the point P_0 is a solution of system (6). Then the implicit function

theorem (Theorem 2 of Sec. 3.2) gives the condition for the existence of a solution in the form: rank $df = m$. Hence, the rank of the Jacobi matrix of f or, what comes to the same thing, the rank of the matrix of the differential df_{P_0} is equal to the dimension of the tangent space $T_{Q_0}(M_2)$. This means that the linear mapping df_{P_0} is an epimorphism. We come therefore to a generalization of Theorem 2 of Sec. 3.2.

Theorem 1. *Let $f\colon M_1 \to M_2$ be a smooth mapping of smooth manifolds, $Q_0 \in M_2$ a regular point of f. Then the inverse image $M_3 = f^{-1}(Q_0)$ is a smooth manifold,* $\dim M_3 = \dim M_1 - \dim M_2$, *and, moreover, some of the local coordinates in M_1 can be chosen as local coordinates in M_3.*

Proof. To prove that M_3 is a manifold, it suffices to apply Theorem 2 of Sec. 3.2 in a neighbourhood of each point $P_0 \in M_3$. We obtain that each point $P_0 \in M_3$ admits a neighbourhood $U \ni P_0$ homeomorphic to a domain in a Euclidean space $\mathbf{R}^{n-m}$, where $n = \dim M_1$, $m = \dim M_2$. Furthermore, some $(n-m)$ of the local coordinates $(x^1, \ldots, x^n)$ of M_1 in a neighbourhood of P_0 can be chosen as local coordinates in U. If these are $(x^{i_1}, \ldots, x^{i_{n-m}})$, the remaining local coordinates (x^j) are expressed on M_3 in terms of smooth functions of $(x^{i_1}, \ldots, x^{i_{n-m}})$. It follows that M_3 is a smooth manifold. Indeed, let $(y^1, \ldots, y^n)$ be another system of local coordinates on M_1 and let $(y^{j_1}, \ldots, y^{j_{n-m}})$ form a local coordinate system on M_3. Then,

$$y^{j_k} = y^{j_k}(x^1, \ldots, x^n)$$

$$= y^{j_k}(x^1(x^{i_1}, \ldots, x^{i_{n-m}}), \ldots, x^n(x^{i_1}, \ldots, x^{i_{n-m}}))$$

is a smooth function. Theorem 1 is proved.

It remains to find an analogue of another property of a linear mapping, the monomorphism property.

Definition 3. Let $f\colon M_1 \to M_2$ be a smooth mapping. The mapping f is called an *immersion* if at each point $P \in M_1$ the differential $df_P\colon T_P(M_1) \to T_{f(P)}(M_2)$ is a monomorphism, i.e. a one-to-one mapping onto its image. If moreover f maps bijectively M_1 onto its image $f(M_1)$ and this image is a closed set, the mapping f is called an *embedding*. The image $f(M_1)$ (as well as M_1) is called in this case a *submanifold* in M_2.

Example 5. The inverse image of a regular point of $f\colon M_1 \to M_2$ is, according to Theorem 1, a submanifold. Indeed, since some of the local coordinates of the ambient manifold M_1 can be taken as a local coordinate system in M_3, the identity mapping $\varphi\colon M_3 \to$

M_1 in the local coordinates takes the form

$$\begin{aligned} x^1 &= x^1, \\ &\cdot \\ &\cdot \\ &\cdot \\ x^{n-m} &= x^{n-m}, \\ x^{n-m+1} &= x^{n-m+1}(x^1, \ldots, x^{n-m}), \\ &\cdot \\ &\cdot \\ &\cdot \\ x^n &= x^n(x^1, \ldots, x^{n-m}). \end{aligned}$$

Therefore, the Jacobi matrix $d\varphi$ includes an identity square matrix. Hence, rank $d\varphi = n - m$, i.e. $d\varphi$ is a monomorphism.

Example 6. Let us consider a mapping $f\colon S^1 \to \mathbf{R}^2$ given by $f(\varphi) = \{\cos\varphi, \sin 2\varphi\}$. The velocity vector $\frac{df}{d\varphi} = \{-\sin\varphi, 2\cos 2\varphi\}$ does not vanish at any point, that is, the rank of the Jacobi matrix

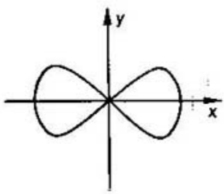

Figure 3.22

is equal to unity. Thus, f is an immersion (see Fig. 3.22). The curve in Fig. 3.22 is a Lissajous' figure; it can be obtained on an oscilloscope if sinusoidal signals are applied to the horizontal and vertical deflecting plates.

3.4.3. SARD'S THEOREM

We have shown in Sec. 3.4.2 how the local properties of a smooth manifold can be deduced from the properties of the differential. Conversely, in certain cases we can find the properties of a differential from the properties of the manifold itself. For example, if $f\colon M_1 \to M_2$ is a smooth homeomorphism, then the differential $df\colon T_P(M_1) \to T_{f(P)}(M_2)$ is, by Lemma 3 of Sec. 3.1, an isomorphism. Consider now a problem which is inverse, to a certain extent,

to that solved in Sec. 3.4.2. Let $f: M_1 \to M_2$ be a smooth mapping of M_1 onto the entire manifold M_2, i.e. $f(M_1) = M_2$. Such a mapping can be considered as an analogue of an epimorphism for a linear mapping. A question then arises: is the differential $df: T_P(M_1) \to T_{f(P)}(M_2)$ an epimorphism? Unfortunately, the answer is no. Consider the following example. Let $M_1 = M_2 = \mathbf{R}^1$, $f: \mathbf{R}^1 \to \mathbf{R}^1$, $f(x) = x^3$, $x \in \mathbf{R}^1$. In this case f is a smooth mapping and $f(\mathbf{R}^1) = \mathbf{R}^1$. But at point $x = 0$ the differential df equals zero and is not therefore an epimorphism. At other points the differential $df = 3x^2dx$ is an epimorphism. This example suggests the general answer to the question. We shall formulate it as the following statement.

Theorem 2 (Sard's theorem). *Let $f: M_1 \to M_2$ be a smooth mapping of compact manifolds. Then the set G of regular points $Q \in M_2$ of f is open and everywhere dense.*

Before proceeding with the proof of Theorem 2, we shall consider several examples.

Example 7. Let $f: \mathbf{R}^1 \to \mathbf{R}^1$, $f(x) \equiv a = \text{const}$. In this case the differential df is not an epimorphism at any point, but the image $f(\mathbf{R}^1)$ consists of a single point a, i.e. by definition, any point $y \neq a$ is regular (since $f^{-1}(y) = \varnothing$). Hence, the set of regular points is open and everywhere dense.

Example 8. Let $f: \mathbf{R}^1 \to \mathbf{R}^1$ be a finite smooth function and $F = \{x: f'(x) = 0\}$. The set F is closed. The image $f(F)$ is a compact set and consists of all non-regular points. We shall demonstrate that $f(F)$ is nowhere dense. If this is not so, we can find an interior point $y \in f(F)$, i.e. y is contained in $f(F)$ together with a neighbourhood $y \in U \subset f(F)$. Since f is finite, the image $f(F)$ lies in the image of the interval $f([a, b])$. In other words, it is sufficient to prove that the image $f(F')$, $F' = F \cap [a, b]$ is nowhere dense. Let $V \supset F'$ be a neighbourhood of the set F', $V \subset (-2a, 2a)$. Then $f(V)$ contains U and hence there exists a point $x \in V$ such that $f'(x)$ exceeds, by modulus, the number $\varepsilon = \operatorname{diam} U/4a$. Diminishing the neighbourhood V, we obtain a sequence of points $x_n \to F'$. We may assume, without loss of generality, that $x_n \to x_0 \in F'$. Then, $f'(x_n) \to f'(x_0)$, i.e. $|f'(x_0)| \geqslant \varepsilon$, which contradicts the condition $f'(x_0) = 0$ at point $x_0 \in F' \subset F$.

It is more convenient to formulate Theorem 2 in more general terms: *if $F \subset M_1$ is a compact set consisting of non-regular points, then $f(F)$ is nowhere dense.* Demonstrate that it is sufficient to prove Theorem 2 for the case where M_1 is a neighbourhood of a closed disk in a Euclidean space. Indeed, cover M_1 with a finite atlas U_α and choose a covering $V_\alpha \subset \overline{V}_\alpha \subset U_\alpha$ such that V_α is homeomorphic to a disk in a Euclidean space. Let $G_\alpha \subset M_2$ be a set of regular points for the mapping f on $\overline{V}_\alpha$. Then the intersection $G = \bigcap G_\alpha$ consists of regular points of the entire mapping f. If G_α are open sets which are everywhere dense, then G is also open and everywhere dense.

Choose a sufficiently fine atlas U_α such that the image $f(U_\alpha)$ lies in a single chart W_β of M_2. Then Theorem 2 can be proved for regular points of the mapping $f|_{U_\alpha}: U_\alpha \to W_\beta$ on $\bar{V}_\alpha$. Indeed, if $G \subset W_\beta$ is the set of regular points of $f|_{U_\alpha}$, then $G \cup (M_2 \setminus f(\bar{V}_\alpha))$ is the set of regular points of the mapping $f: U_\alpha \to M_2$ on $\bar{V}_\alpha$. Thus, let U be a neighbourhood of a disk D^n in $\mathbf{R}^n$ and let $f: U \to \mathbf{R}^m$ be a smooth mapping. We shall demonstrate that the set of points $y \in \mathbf{R}^m$, for which $D^n \cap f^{-1}(y)$ consists of regular points, is open and everywhere dense.

Lemma 3. *Theorem 2 is valid for $m = 1$.*

Proof. Let $F \subset D^n$ be a set of non-regular points of a function f. Then $f(F)$ is compact and contains all non-regular points of f. Demonstrate that $\mathbf{R}^1 \setminus f(F)$ is everywhere dense. If this is not the case, there exists an interval $V \subset f(F)$. Fix $k > n$ and consider the set F_k of those points for which all partial derivatives of f of order up to k inclusive are zero. Expanding f in a Taylor series in a neighbourhood of an arbitrary point $y \in F_k$, we obtain $|f(y) - f(x)| < C|x - y|^k$, where C does not depend on the choice of $y \in F_k$ and $x \in D^n$. This means that if F_k is covered with cubes of side $1/N$ (the number of these cubes does not exceed N^n), the image $f(F_k)$ will be covered with intervals, the length of each interval not exceeding $2\sqrt{n^k}\, C/N^k$. Hence, the sum of the lengths of all the intervals does not exceed $2\sqrt{n^k} C/N^{k-n}$ and vanishes for $N \to \infty$, that is, $f(F_k)$ is nowhere dense.

The remaining part of the set F, i.e. $F \setminus F_k$, can be represented as a union of a finite family of subsets, each lying on a submanifold defined by one of the equations

$$\frac{\partial^l f}{\partial x_1^{l_1} \ldots \partial x_n^{l_n}} = 0, \quad l_1 + \ldots + l_n = l < k.$$

Indeed, let $F_{l_1, \ldots, l_n}$ be the set of those points in F at which

$$\frac{\partial^l f}{\partial x_1^{l_1} \ldots \partial x_n^{l_n}} = 0, \quad \operatorname{grad} \frac{\partial^l f}{\partial x_1^{l_1} \ldots \partial x_n^{l_n}} \neq 0. \tag{6}$$

Obviously, $F \setminus F_k = \bigcup\limits_{l_1 + \ldots + l_n < k} F_{l_1, \ldots, l_n}$. On the other hand, the set $F_{l_1 \ldots l_n}$ lies on the submanifold $M_{l_1 \ldots l_n}$ of those points where conditions (6) are satisfied. The dimension of $M_{l_1 \ldots l_n}$ is less than n, and we may conclude, by induction, that Lemma 3 is valid for $M_{l_1 \ldots l_n}$. Thus, $f(F_k)$ does not cover the interval V. Hence, there exists a neighbourhood $U_k \supset F_k$ such that $\overline{f(U_k)}$ does not cover V. Let $l_1 + \ldots + l_n = k - 1$. Then $F_{l_1 \ldots l_n} \setminus U_k$ is a compact set on $M_{l_1 \ldots l_n}$, so that $f(F_{l_1 \ldots l_n} \setminus U_k)$ does not cover $V \setminus \overline{f(U_k)}$, i.e. $f(F_k \cap F_{l_1 \ldots l_n})$ does not cover V. Hence, there

exists a neighbourhood $U_s \supset F_k \cup \bigcup_{s<l_1+\ldots+l_n<k} F_{l_1\ldots l_n}$ such that $f(\overline{U_s})$ does not cover the interval V. Thus, for $l_1+\ldots+l_n=s$ the sets $F_{l_1\ldots l_n}\setminus U_s$ do not cover the remainder $V\setminus f(F_k \cup \bigcup_{s<l_1+\ldots+l_n<k} F_{l_1\ldots l_n})$ under the mapping f, and therefore $f(F_k \cup \bigcup_{s\leq l_1+\ldots+l_n<k} F_{l_1\ldots l_n})$ does not cover V. After a finite number of steps we find that $f(F)$ does not cover V.

We now apply Lemma 3 to prove, by induction on m, Theorem 2 for a system of functions $f\colon U\to \mathbf{R}^m$, $D^n\subset U$, $f(P)=(f^1(P), \ldots, f^m(P))$. Since f^1 is a smooth function, we find by Lemma 3 that the set G_1 of regular values of f^1 is open and everywhere dense in $\mathbf{R}^1$. Let $y_0^1\in G_1$, $N=(f^1)^{-1}(y_0^1)$. By Theorem 1, the set N is a smooth submanifold mapped by f into a hyperplane $\mathbf{R}^{m-1}$. Then the point $(y_0^2, \ldots, y_0^m)$ is regular for the mapping $f|_N$ if and only if the point $(y_0^1, \ldots, y_0^n)$ is regular for f. By induction, the set of points $(y_0^2, \ldots, y_0^m)$ which are regular for the mapping $f|_N$ is everywhere dense in $\mathbf{R}^{m-1}$; hence, the set of regular points for f is also everywhere dense in $\mathbf{R}^m$. In order to demonstrate that the set of regular points is open we note that the inverse image of a regular point $D^n \cap f^{-1}(y_0^1, \ldots, y_0^n)$ is a compact and the minor of the matrix of the differential df is non-zero at each point of the compact. Therefore, for any sufficiently close point $(y_1^1, \ldots, y_1^n)$ the inverse image $D^n\cap f^{-1}(y_0^1, \ldots, y_0^n)$ lies in quite a small neighbourhood of $D^n\cap f^{-1}(y_0^1, \ldots, y_0^n)$, that is, the same minors are non-zero. This means that the set of regular points of f is open. Theorem 2 is proved.

As an application of Sard's theorem, we shall consider a smooth mapping $f\colon M_1\to M_2$ for $\dim M_1<\dim M_2$. Then, none of the points $P\in M_1$ can be regular, and this means that the image $f(M_1)$ is nowhere dense in M_2. In particular, the image of f does not cover M_2.

Remark. Sard's theorem can be generalized to non-compact separable manifolds. In this case the set of regular points need not necessarily be an open set, but it should only be the intersection of a countable number of open sets which are everywhere dense. Such sets are called G_δ-sets. It is known from general topology that the intersection of a countable number of sets, which are open and everywhere dense in $\mathbf{R}^n$, is always non-empty and everywhere dense. Hence, in the case of non-compact manifolds the set of regular points is non-empty and everywhere dense.

3.4.4. EMBEDDING OF MANIFOLDS IN A EUCLIDEAN SPACE

In conclusion let us state the Whitney theorem which asserts that any compact manifold is a submanifold of a Euclidean space of a sufficiently large dimension.

Theorem 3. *Let M be a smooth compact manifold. Then for a proper dimension N there exists the embedding $\varphi: M \to \mathbf{R}^N$.*

Proof. Let $\{U_\alpha\}_{\alpha=1}^M$ be a finite atlas and $(x_\alpha^1, \ldots, x_\alpha^n)$ a local coordinate system in the chart U_α. We may assume without loss of generality that the U_α are homeomorphic to a ball $D^n \subset \mathbf{R}^n$ of radius 1 and the coordinates $(x_\alpha^1, \ldots, x_\alpha^n)$ perform a homeomorphism φ_α of U_α onto the ball D^n. We may also assume that D^n lies in $\mathbf{R}^n$ and does not contain the origin (this can be achieved by a translation in $\mathbf{R}^n$). Further, let $D_1^n \subset D^n$ be a ball of smaller radius with the same centre as D^n, let the manifold M be covered with open sets $U'_\alpha = \varphi_\alpha^{-1}(D_1^n) \subset U_\alpha$, and let f be a smooth function in $\mathbf{R}^n$ such that it is identically unity on D_1^n and supp $f \subset D^N$. We put then

$$y_\alpha^k(P) = \begin{cases} f(\varphi_\alpha(P))\, x_\alpha^k(P) & \text{if} \quad P \in U_\alpha, \\ 0 & \text{if} \quad P \notin U_\alpha. \end{cases}$$

Obviously, $y_\alpha^k(P) = x_\alpha^k(P)$ if $P \in U'_\alpha$. We have obtained a system of $N = n \cdot M$ smooth functions, $\{y_\alpha^k(P)\}$. This system defines the mapping g of M into a Euclidean space $\mathbf{R}^N$, $g(P) = \{y_\alpha^k(P)\} \in \mathbf{R}^N$. We now demonstrate that the differential of g is of rank n at each point. Let $P \in M$ be an arbitrary point, $U'_\alpha \ni P$, and $(x_\alpha^1, \ldots, x_\alpha^n)$ a local coordinate system. The Jacobi matrix of g at point P in the local coordinate system $(x_\alpha^1, \ldots, x_\alpha^n)$ consists of partial derivatives $\left\{\frac{\partial y_\beta^k}{\partial x_\alpha^j}(P)\right\}$. In particular for $\beta = \alpha$,

$$\frac{\partial y_\beta^k}{\partial x_\alpha^j}(P) = \frac{\partial x_\alpha^k}{\partial x_\alpha^j}(P) = \delta_j^k,$$

i.e. the Jacobi matrix of g contains an identity matrix of order n. Hence, rank $dg = n$.

The mapping g is thus an immersion. In order that g be an embedding, it is necessary that distinct points P and Q be mapped into distinct points $g(P)$ and $g(Q)$. Construct a new mapping $\bar{g}(P) = \{y_\alpha^k(P),\ f(\varphi_\alpha(P))\} \in \mathbf{R}^{N+M}$. Owing to the same arguments as for g, this mapping is an immersion. Let $P \neq Q$ be two points on a manifold. Consider a number α such that $f(\varphi_\alpha(P)) = 1$. If $f(\varphi_\alpha(Q)) < 1$, then $\bar{g}(P) \neq \bar{g}(Q)$; if $f(\varphi_\alpha(Q)) = 1$, then $y_\alpha^k(P) = x_\alpha^k(P)$, $y_\alpha^k(Q) = x_\alpha^k(Q)$, and for a certain number k we have $x_\alpha^k(P) \neq x_\alpha^k(Q)$, i.e. $\bar{g}(P) \neq \bar{g}(Q)$. Thus, the mapping $g: M \to \mathbf{R}^{N+M}$ is a one-to-one immersion, i.e. an embedding. Theorem 3 is proved.

Hence, any smooth compact manifold M may be considered as embedded (in the form of a submanifold) in a Euclidean space $\mathbf{R}^N$

of a rather large dimension N. In practice, however, the dimension of $\mathbf{R}^N$ can be reduced significantly. For example, a sphere S^n can be embedded in $\mathbf{R}^{n+1}$ and a torus T^n in $\mathbf{R}^{2n}$. The projective plane $\mathbf{RP}^2$ cannot be embedded in $\mathbf{R}^3$, but it can be embedded in $\mathbf{R}^5$. Indeed, let $(x_1 : x_2 : x_3)$ be homogeneous coordinates of a point P in $\mathbf{RP}^2$. Putting

$$y^1 = \frac{x_1^2}{x_1^2+x_2^2+x_3^2}, \quad y^2 = \frac{x_2^2}{x_1^2+x_2^2+x_3^2}, \quad y^3 = \frac{x_3^2}{x_1^2+x_2^2+x_3^2},$$
$$y_4 = \frac{x_1 x_2}{x_1^2+x_2^2+x_3^2}, \quad y^5 = \frac{x_2 x_3}{x_1^2+x_2^2+x_3^2}, \quad y^6 = \frac{x_3 x_1}{x_1^2+x_2^2+x_3^2},$$

we obtain the mapping $g\colon \mathbf{RP}^2 \to \mathbf{R}^6$, $g(P) = g(x_1 : x_2 : x_3) = (y^1, y^2, y^3, y^4, y^5, y^6)$. It appears however that the image of g lies in the linear subspace $\mathbf{R}^5 \subset \mathbf{R}^6$ defined by the equation $y^1 + y^2 + y^3 = 1$. Verify that g is an immersion, i.e. the differential dg is a monomorphism at each point $P \in \mathbf{RP}^2$. In other words, we have to prove that in any local coordinate system the rank of the Jacobi matrix of g is equal to 2. All the coordinate functions of g are symmetric with respect to the permutation of homogeneous coordinates $(x_1 : x_2 : x_3)$. Without loss of generality we can therefore assume that $x_1 \neq 0$ in a neighbourhood of point $P_0 \in \mathbf{RP}^2$, so that the remaining coordinates x_2, x_3 (for $x_1 = 1$) can be chosen as a local coordinate system. The Jacobi matrix of g is

$$dg = \begin{pmatrix} -2x_2 & -2x_3 \\ 2x_2(1+x_3^2) & -2x_2^2 x_3 \\ -2x_3^2 x_2 & 2x_3(1+x_2^2) \\ (1-x_2^2+x_3^2) & -2x_2 x_3 \\ x_3(1-x_2^2+x_3^2) & x_2(1+x_2^2-x_3^2) \\ -2x_2 x_3 & (1+x_2^2-x_3^2) \end{pmatrix}$$

to within a proportionality factor. If $x_2 x_3 \neq 0$, the minor of the first two rows does not vanish. If $x_2 = 0$, $x_3 = 0$, the minor of the fourth and sixth rows is not zero, and if $x_2 = 0$, $x_3 \neq 0$, the minor of the first and fourth rows is also non-zero. Thus, $\operatorname{rank} dg = 2$, i.e. g is an immersion. Demonstrate that g is a one-to-one mapping. Without loss of generality we may assume that the homogeneous coordinates are so chosen that $x_1^2 + x_2^2 + x_3^2 = 1$. Let $P = (x_1 : x_2 : x_3)$, $Q = (x_1' : x_2' : x_3')$. Then

$$g(P) = (x_1^2,\ x_2^2,\ x_3^2,\ x_1 x_2,\ x_2 x_3,\ x_3 x_1),$$
$$g(Q) = (x_1'^2,\ x_2'^2,\ x_3'^2,\ x_1' x_2',\ x_2' x_3',\ x_3' x_1').$$

If $g(P) = g(Q)$, then $x_1^2 = x_1'^2$. If $x_1 \neq 0$, then $x_1 = \pm x_1'$, and we can take $x_1 = x_1'$, since homogeneous coordinates can be multiplied by ± 1. Since the fourth and sixth coordinates are equal, we obtain

$x_2 = x_2'$, $x_3 = x_3'$, i.e. $P = Q$. Since all the coordinates x_1, x_2, and x_3 are equivalent, we always have $P = Q$. Thus, g is an embedding of a projective plane $\mathbf{RP}^2$ in a five-dimensional Euclidean space $\mathbf{R}^5$.

An analogous statement holds true for any compact manifold M.

Theorem 4. *Let M be a smooth compact manifold of* $\dim M = n$. *Then the embedding* $\varphi\colon M \to \mathbf{R}^{(2n+1)}$ *does exist.*

Proof. We shall use Theorem 2 and try to reduce the dimension of $\mathbf{R}^N$. Let $\mathbf{e} \in \mathbf{R}^N$ be a non-zero vector and $p_\mathbf{e}: \mathbf{R}^N \to \mathbf{R}^{N-1}$ an orthogonal projection along $\mathbf{e}$ onto a subspace orthogonal to $\mathbf{e}$. Sometimes, the composition $M \xrightarrow{\varphi} \mathbf{R}^N \xrightarrow{p_\mathbf{e}} \mathbf{R}^{N-1}$ remains an embedding. Analyse the conditions under which the composition $p_\mathbf{e}\varphi$ is an embedding. We have to verify two conditions: (a) that the differential is a monomorphism and (b) that the mapping is one-to-one.

We first consider condition (a). Let $P \in M$, $V_P = d\varphi_P(T_PM) \subset \mathbf{R}^N$. In order that the differential of the composition $d(p_\mathbf{e}\varphi)$ at a point P be a monomorphism, it is necessary and sufficient that the projection $p_\mathbf{e}$ should map the subspace V_P monomorphically into $\mathbf{R}^{N-1}$, which is equivalent to $\mathbf{e} \notin V_P$. Fix a local coordinate system $(x^1, \ldots, x^n)$ in a neighbourhood U of point P and a basis in $\mathbf{R}^n$. Construct the mapping $h\colon U \times \mathbf{R}^n \to \mathbf{RP}^{N-1}$, assuming $h(x^1, \ldots, x^n, \xi^1, \ldots, \xi^n) = (\zeta^1 : \zeta^2 : \ldots : \zeta^N)$, where $\zeta^k = \sum_{\alpha=1}^{n} \frac{\partial y^k}{\partial x^\alpha} \xi^\alpha$ and $(y^1, \ldots, y^N) = \varphi(x^1, \ldots, x^n)$. The mapping h is a smooth mapping of the $2n$-dimensional manifold $U \times \mathbf{R}^n$ into the projective space $\mathbf{RP}^{N-1}$. Then the condition $\mathbf{e} \notin V_P$ exactly means that the straight line generated by $\mathbf{e}$, being a point in $\mathbf{RP}^{N-1}$, does not belong to the image of h. By Sard's theorem, for $2n < N - 1$ the set of such points is open and everywhere dense in the entire space $\mathbf{RP}^{N-1}$. Considering a finite atlas in M, we find that for an open dense set G in $\mathbf{RP}^{N-1}$ the vector $\mathbf{e}$ generating points in G satisfy $\mathbf{e} \notin V_P$ for any point $P \in M$. This means that the set of $\mathbf{e}$ such that the projection $p_\mathbf{e}$ embeds $\varphi(M)$ in $\mathbf{R}^{N-1}$ is open and dense. Let us now turn to condition (b). The absence of bijectiveness means that $\mathbf{e}$ is parallel to a straight line through a pair of points $P \neq Q$ on $\varphi(M)$. Just like in the case of condition (a), we shall consider the mapping h': $(M \times M \setminus \Delta) \to \mathbf{RP}^{N-1}$ which to a pair of points $P \neq Q$ associates a straight line through $\varphi(P)$ and $\varphi(Q)$. According to Sard's theorem, for $2n < N - 1$ the set of points not contained in the image of h' is an open dense set G'. The vector $\mathbf{e}$ should therefore be so chosen that the straight line it generates lie in $G \cap G' \neq \varnothing$. We have demonstrated that if $2n < N - 1$, there exists a projection p such that the composition $p_\mathbf{e}\varphi$ is an embedding. Hence, step by step we can reduce the dimension of the surrounding Euclidean space $\mathbf{R}^N$ unless the equality $2n = N - 1$ is satisfied, whence $N = 2n + 1$.

This is the minimal dimension of the Euclidean space $\mathbf{R}^{(2n+1)}$ which admits an embedding of any n-dimensional compact manifold M. Theorem 4 is proved.

Now we are able to give the general concept of a Riemannian metric on a manifold (particular cases of this concept have been considered above).

The *Riemannian metric* on a smooth manifold M is the family of positive definite scalar products defined in each tangent space $T_P(M)$. If a coordinate system $(x^1, \ldots, x^n)$ is valid in a neighbourhood U of a point P, a coordinate system $(\xi^1, \ldots, \xi^n)$ is also valid in $T_P(M)$. The scalar product in $(\xi^1, \ldots, \xi^n)$ is defined by a point-dependent non-singular symmetric matrix $G = (g_{ij})$. In a new coordinate system $(y^1, \ldots, y^n)$ the matrix takes the form

$$g'_{ij}(P) = \sum_{\alpha, \beta=1}^{n} g_{\alpha\beta}(P) \frac{\partial x^\alpha}{\partial y^i}(P) \frac{\partial x^\beta}{\partial x^j}(P),$$

where $G' = (g'_{ij}(P))$ is the scalar product matrix in the new coordinate system $(y^1, \ldots, y^n)$.

Definition 4. The *Riemannian metric* on a smooth manifold M is the family of non-degenerate positive definite scalar products in each tangent space $T_P(M)$, $P \in M$, such that in a local coordinate system the scalar product matrix is a smooth function of the local coordinates.

The metric can be defined as a correspondence which to every coordinate system $(x^1_\alpha, \ldots, x^n_\alpha)$ in the chart U_α associates a matrix-valued smooth function in the chart U_α: $G^\alpha(P) = (g^\alpha_{ij}(P))$ such that: (a) the matrix $G^\alpha(P)$ is symmetric and positive definite, and (b) the coefficients $g^\alpha_{ij}(P)$ are transformed as

$$g^\alpha_{ij}(P) = \sum_{k, l=1}^{n} g^\beta_{kl}(P) \frac{\partial x^k_\beta}{\partial x^i_\alpha}(P) \frac{\partial x^l_\beta}{\partial x^j_\alpha}(P)$$

for points $P \in U_\alpha \cap U_\beta$. If $\xi = (\xi^1, \ldots, \xi^n)$ and $\eta = (\eta^1, \ldots, \eta^n)$ are two tangent vectors, the scalar product ξ, η is written as

$$(\xi, \eta) = \sum_{i, j=1}^{n} g^\alpha_{ij}(P) \xi^i_\alpha \eta^j_\alpha.$$

Problems

1. Prove, using Sard's theorem, that any smooth compact manifold M of dimension n can be embedded in $\mathbf{R}^{2n+1}$ and immersed in $\mathbf{R}^{2n}$.

Hint: consider the projections $\mathbf{R}^N \to \mathbf{R}^{N-1}$ and choose an appropriate one.

2. Construct an embedding of a torus T^n in $\mathbf{R}^{n+1}$.

Hint: represent T^n as a hypersurface of rotation T^{n-1} about an axis.

3. Construct the embedding of $S^2 \times S^2$ in $\mathbf{R}^5$.

Hint: in $\mathbf{R}^5$ find an open domain homeomorphic to $S^2 \times \mathbf{R}^3$.

4. Prove, using Sard's and Whitney's theorems, that any smooth mapping $f: M^n \to \mathbf{R}^{2n+1}$ of a compact manifold M^n can be approximated as close as possible by an embedding.

Hint: complete f to the embedding $\mathbf{R}^N = \mathbf{R}^{N-(2n+1)}\mathbf{R}^{(2n+1)}$ and consider the projection along a subspace close to $\mathbf{R}^{N-(2n+1)}$.

Chapter 4

Smooth Manifolds (Examples)

4.1. THE THEORY OF CURVES ON A PLANE AND IN A THREE-DIMENSIONAL SPACE

4.1.1. THE THEORY OF CURVES ON A PLANE. FRENET FORMULAS

We shall consider a Euclidean plane $\mathbf{R}^2$ referred to Cartesian coordinates (x, y). A smooth curve $\gamma(t)$ on $\mathbf{R}^2$ is defined by two functions $x(t)$ and $y(t)$, i.e. we deal with the radius vector $\mathbf{r}(t) = ((x(t), y(t))$ of a smooth curve $\gamma(t)$ emerging from the origin, point 0. Recall that the velocity vector $\mathbf{v}(t)$ of a curve $\gamma(t)$ at a point t is the vector with the coordinates $\left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}\right)$. The straight line defined by this vector is tangent to the curve $\gamma(t)$ at a point t. Here we assume, of course, that $\mathbf{v}(t) \neq 0$. This assumption is supposed to hold throughout this section (at the points where the velocity vector vanishes a smooth curve may suffer a cusp; the examples have been given in Chapter 1). The derivative of a radius vector is sometimes denoted by

$$\dot{\mathbf{r}}(t) = \mathbf{v}(t) = (\dot{x}(t), \dot{y}(t))$$

and the symbol $\left|\frac{d\mathbf{r}(t)}{dt}\right| = |\mathbf{v}(t)|$ will stand for the magnitude of velocity vector (in the Euclidean metric). Let s denote the length of a curve from a fixed point to a variable point; since the length of an arc monotonically increases as the variable point moves along the curve, we may take this length as a parameter along the curve. This parameter is called a *natural parameter*, and the equation of the curve $\mathbf{r}(s) = (x(s), y(s))$ written as a vector function of s is called *natural parametrization of a curve.*

Lemma 1. *The magnitude of the velocity vector of a curve written in the natural parameter is constant and equal to unity.*

Proof. We have for the length of arc $l(\gamma)_a^b = \int\limits_a^b \left|\frac{d\mathbf{r}(t)}{dt}\right| dt$. It follows that the identity $dl = dt\left|\frac{d\mathbf{r}(t)}{dt}\right|$ holds at each point which is what was required.

Thus, the motion along a curve referred to the natural parameter occurs at a constant velocity (only the magnitude of the velocity

vector is constant, but not its direction). *Corollary*: at each point of the curve the velocity vector $\mathbf{v}(s)$ is non-zero. Here we have used the fact that the relation $\frac{d\mathbf{r}(t)}{dt} \neq 0$ is satisfied for the initial parameter t. With each point on a curve we can associate, besides the velocity vector, another vector smoothly dependent on a point, $\mathbf{q}(s) = \frac{d\mathbf{v}(s)}{ds}$. This vector is orthogonal to the velocity vector, which is a direct consequence of the following lemma.

Lemma 2. *Let* $\mathbf{p}(t)$ *be a vector function such that* $|\mathbf{p}(t)| \equiv 1$. *Then the vector* $\frac{d\mathbf{p}(t)}{dt}$ *is orthogonal to* $\mathbf{p}(t)$ $\left(\textit{for any } t \textit{ such that } \frac{d\mathbf{p}(t)}{dt} \neq 0\right)$.

Proof. The condition of the lemma implies that $(\mathbf{p}, \mathbf{p}) \equiv 1$. Differentiating this identity with respect to t, we obtain

$$\left(\frac{d\mathbf{p}}{dt}, \mathbf{p}\right) + \left(\mathbf{p}, \frac{d\mathbf{p}}{dt}\right) \equiv 0, \quad \text{i.e.} \quad \left(\frac{d\mathbf{p}}{dt}, \mathbf{p}\right) \equiv 0,$$

The lemma is proved.

Thus, at each point of a smooth curve $\gamma(s)$ referred to its natural parameter there exist two orthogonal vectors: the velocity vector and acceleration vector. The latter need not have the unit length. It is convenient to introduce the unit vector $\mathbf{n}(s) = \frac{d\mathbf{v}(s)}{ds} \Big/ \left|\frac{d\mathbf{v}(s)}{ds}\right|$. Hence, as the parameter s changes, there arises along the curve a smooth coordinate frame field, i.e. the family of two-dimensional frames $(\mathbf{v}(s), \mathbf{n}(s))$. The vector $\mathbf{n}(s)$ is called a *normal vector* to a curve at point s. Each frame, after being translated to the origin, uniquely defines a rotation of a plane about the point 0. The frame field along a curve defines therefore a smooth mapping of $\gamma(s)$ into a group of orthogonal matrices, i.e. a group of rotations of a plane. In other words, we can say that each curve $\gamma(s)$ generates a smooth curve whose points are represented as orthogonal 2×2 matrices. We shall study the properties of this curve for a multi-dimensional case, i.e. for a smooth curve in a multi-dimensional Euclidean space (see below).

Definition 1. Let a smooth curve be referred to the natural parameter. The *curvature of a curve* at a point s is the number $k(s) = \left|\frac{d^2\mathbf{r}(s)}{ds^2}\right|$.

Since the curvature (by definition) is the magnitude of the acceleration vector, we have $\frac{d^2\mathbf{r}(s)}{ds^2} = k(s)\,\mathbf{n}(s)$, where $\mathbf{n}(s)$ is the normal vector at the point s and

$$k(s) = \sqrt{\left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2}.$$

If e_1 and e_2 are orthogonal unit vectors defining Cartesian coordinates on the plane $R^2(x, y)$, the normal vector can be written explicitly as

$$\mathbf{n}(s) = \left(\frac{d^2x}{\partial s^2} + \frac{d^2y}{ds^2}\right)^{-\frac{1}{2}} \cdot \left(\frac{d^2x}{ds^2}\,\mathbf{e}_1 + \frac{d^2y}{ds^2}\,\mathbf{e}_2\right).$$

Definition 2. The *curvature radius* of a smooth curve at a point s is the number $\mathbf{R}(s) = 1/k(s)$.

Before proceeding further, we shall consider simple examples which underlie the choice of the terms "curvature" and "curvature radius". The simplest curve on a plane is a straight line defined parametrically as a linear vector function: $x(s) = x(0) + \alpha s$, $y(s) = y(0) + \beta s$, where s is the natural parameter. This imposes restrictions on the numbers α and β: obviously, the equality $\sqrt{\alpha^2 + \beta^2} = 1$ must be satisfied, since $\mathbf{v} = (\alpha, \beta)$ and $|\mathbf{v}(s)| \equiv 1$. Then the acceleration vector $\frac{d\mathbf{v}(s)}{ds}$ is identically zero, and the curvature of a straight line is also zero. Hence, the curvature radius of a straight line is equal to infinity.

Let us consider a circle of radius R on a plane. The parametric equations of the circle are

$$x(s) = x(0) + R\cos\left(\frac{s}{R}\right), \quad y(s) = y(0) + R\sin\left(\frac{s}{R}\right).$$

The parameter s varies from 0 to 2π. The velocity vector is

$$\mathbf{v}(s) = \left(-\sin\left(\frac{s}{R}\right),\ \cos\left(\frac{s}{R}\right)\right).$$

Hence,

$$\frac{d\mathbf{v}(s)}{ds} = \left(-\frac{1}{R}\cos\left(\frac{s}{R}\right), \quad -\frac{1}{R}\sin\left(\frac{s}{R}\right)\right).$$

Thus, the curvature of a circle is constant and equal to $1/R$ and the curvature radius is $1/k = R$.

In many specific problems it appears, however, that the equations of a curve are referred to an arbitrary parameter t, rather than to the natural parameter. It would be useful therefore to have formulas for calculating the curvature of a curve referred to an arbitrary parameter.

Theorem 1. *Let a smooth curve $\gamma(t)$ be referred to an arbitrary parameter t, not necessarily natural, and let the velocity vector $\mathbf{v}(t)$ be nonzero at point t. Then the formula*

$$k(s) = \left|\frac{d^2\mathbf{r}(s)}{ds^2}\right| = \frac{|x''y' - y''x'|}{[(x')^2 + (y')^2]^{3/2}}$$

is valid, where x', x'', . . . denote derivatives with respect to t.

Proof. Let $\mathbf{r}(t) = (x(t), y(t))$ be the parametric form of $\gamma(t)$ and $\mathbf{v}(t) = (x'(t), y'(t))$ the velocity vector. If s is the natural parameter, we have for an arbitrary vector function $\mathbf{q}(t)$

$$\frac{d}{ds}\mathbf{q}(t) = \frac{d\mathbf{q}(t)}{dt}\cdot\frac{dt}{ds}.$$

Take for $\mathbf{q}$

$$\mathbf{q}(t) = \mathbf{v}(t)/|\mathbf{v}(t)| = \frac{d\mathbf{r}(s)}{ds}.$$

From the definition of curvature we obtain

$$k = \left|\frac{d^2\mathbf{r}}{ds^2}\right| = \left|\frac{d}{ds}\left(\frac{d\mathbf{r}}{ds}\right)\right| = \left|\frac{d}{ds}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right)\right|$$

and further

$$\frac{d}{ds}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right) = \frac{dt}{ds}\cdot\frac{d}{dt}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right).$$

Let us find $\frac{dt}{ds}$. Since $ds = \left|\frac{d\mathbf{r}(t)}{dt}\right|dt$, we have

$$\frac{dt}{ds} = \frac{1}{\left|\frac{d\mathbf{r}(t)}{dt}\right|} = \frac{1}{|\mathbf{v}(t)|},$$

whence

$$\begin{aligned}
k &= \frac{1}{|\mathbf{v}(t)|}\cdot\frac{d}{dt}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right)\\
&= \frac{1}{|\mathbf{v}(t)|^2}\cdot\left(\frac{d\mathbf{v}(t)}{dt} - \frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\cdot\frac{d}{dt}|\mathbf{v}(t)|\right)\\
&= \frac{1}{|\mathbf{r}'|^2}\left(\mathbf{r}'' - \frac{\mathbf{r}'}{|\mathbf{r}'|}\cdot\frac{d}{dt}|\mathbf{r}'|\right)\\
&= \frac{1}{|\mathbf{r}'|^2}\cdot\left(\mathbf{r}'' - \frac{\mathbf{r}'}{2|\mathbf{r}'|^2}\frac{d}{dt}|\mathbf{r}'|^2\right).
\end{aligned}$$

Since

$$\frac{d}{dt}|\mathbf{r}'|^2 = \frac{d}{dt}(\mathbf{r}', \mathbf{r}') = 2(\mathbf{r}', \mathbf{r}''),$$

we have

$$k = \left|\frac{d^2\mathbf{r}(s)}{ds^2}\right| = \frac{1}{|\mathbf{r}'|^2}\cdot\left|\mathbf{r}'' - \mathbf{r}'\cdot\frac{(\mathbf{r}', \mathbf{r}'')}{|\mathbf{r}'|^2}\right|,$$

or in greater detail

$$\begin{aligned}
&\frac{1}{|\mathbf{r}'|^2}\left(\mathbf{r}'' - \mathbf{r}'\cdot\frac{(\mathbf{r}', \mathbf{r}'')}{|\mathbf{r}'|^2}\right)\\
&= \frac{1}{|r'|^2}\left(x'' - x'\cdot\frac{x'x'' + y'y''}{(x')^2 + (y')^2}\right)\mathbf{e}_1 + \frac{1}{|\mathbf{r}'|^2}\left(y'' - y'\cdot\frac{x'x'' + y'y''}{(x')^2 + (y')^2}\right)\mathbf{e}_2\\
&= \frac{1}{|\mathbf{r}'|^2}\left(\frac{x''(y')^2 - x'y'y''}{(x')^2 + (y')^2}\mathbf{e}_1 + \frac{1}{|\mathbf{r}'|^2}\cdot\left(\frac{y''(x')^2 - y'x'x''}{(x')^2 + (y')^2}\right)\mathbf{e}_2.\right.
\end{aligned}$$

Hence,

$$k^2 = ((x')^2 + (y')^2)^{-4} \cdot ((y')^2 (x''y' - x'y'')^2 + (x')^2 (y''x' - y'x'')^2) = \frac{(x''y' - y''x')^2}{((x')^2 + (y')^2)^3}.$$

and $k = \frac{|x''y' - x'y''|}{((x')^2 + (y')^2)^{3/2}}$. The theorem is proved.

Let us return to the motion of the frame $(\mathbf{v}(s), \mathbf{n}(s))$ when the parameter s varies. It appears that the derivatives of the frame vectors satisfy simple relations, called the *Frenet formulas*.

Theorem 2. *If a smooth curve is referred to the natural parameter, the relations*

$$\frac{d\mathbf{v}(s)}{ds} = k(s)\,\mathbf{n}(s), \quad \frac{d\mathbf{n}(s)}{ds} = -k(s)\,\mathbf{v}(s),$$

are satisfied.

Proof. The first Frenet formula directly follows from the definition of curvature $k(s)$. It remains to verify the second formula. Consider the vector function $\mathbf{n}(s)$. By definition, $(\mathbf{n}(s), \mathbf{n}(s)) = 1$, and according to Lemma 2,

$$\left(\mathbf{n}(s), \frac{d}{ds}\mathbf{n}(s)\right) = 0, \quad \text{i.e.} \quad \frac{d\mathbf{n}(s)}{ds} = \lambda(s)\,\mathbf{v}(s),$$

where λ is a smooth function of s. To find this function, we differentiate the identity $(\mathbf{v}, \mathbf{n}) = 0$ with respect to s to obtain

$$\left(\frac{d\mathbf{v}}{ds}, \mathbf{n}\right) + \left(\mathbf{v}, \frac{d\mathbf{n}}{ds}\right) = 0,$$

whence $k(\mathbf{n}, \mathbf{n}) + (\mathbf{v}, \lambda\mathbf{v}) = 0$, i.e. $k = -\lambda$. The theorem is proved.

The vectors $\mathbf{v}$ and $\mathbf{n}$ can be written as a column to give the following form of the Frenet formulas:

$$\begin{pmatrix} d\mathbf{v}/ds \\ d\mathbf{n}/ds \end{pmatrix} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{n} \end{pmatrix}, \text{ where } X = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix}$$

is a skew-symmetric matrix. This relation has a clear geometric meaning. Let us consider a coordinate frame $\omega(s) = (\mathbf{v}(s), \mathbf{n}(s))$ at a point s and move along the curve $\mathbf{r}(s)$ from s to point $s + \Delta s$ (Fig. 4.1). After translating the frame $\omega(s + \Delta s)$ to the point s we obtain two frames at s, $\omega(s)$ and $\omega(s + \Delta s)$, where the frame $\omega(s + \Delta s)$ is obtained from $\omega(s)$ by rotation through an infinitesimal angle $\Delta\varphi$. Hence, we may assume $\omega(s)$ and $\omega(s + \Delta s)$ to be related by the orthogonal transformation $\omega(s + \Delta s) = A(\Delta s)\,\omega(s)$, where $A(\Delta s) = \begin{pmatrix} \cos\Delta\varphi & \sin\Delta\varphi \\ -\sin\Delta\varphi & \cos\Delta\varphi \end{pmatrix}$. Expanding $\cos\Delta\varphi$ and $\sin\Delta\varphi$ in the small increment $\Delta\varphi$ and neglecting terms of the second order

in $\Delta\varphi$, we get

$$A(\Delta\varphi)=\begin{pmatrix}1&0\\0&1\end{pmatrix}+\begin{pmatrix}0&\Delta\varphi\\-\Delta\varphi&0\end{pmatrix}+\dots,$$

$$\text{i.e. } \omega(s+\Delta s)=\omega(s)+\begin{pmatrix}0&\Delta\varphi\\-\Delta\varphi&0\end{pmatrix}\omega(s)+\dots,$$

$$\text{i.e. } \frac{d}{ds}\omega(s)=\lim_{\Delta s\to 0}\frac{1}{\Delta s}\begin{pmatrix}0&\Delta\varphi\\-\Delta\varphi&0\end{pmatrix}\omega(s)=\begin{pmatrix}0&\frac{d\varphi(s)}{ds}\\ \frac{-d\varphi(s)}{ds}&0\end{pmatrix}\omega(s),$$

where $\varphi(s)$ is the angle of rotation of $\omega(s)$ relative to a fixed coordinate frame on the plane (say, relative to $\omega(0)$). At the same

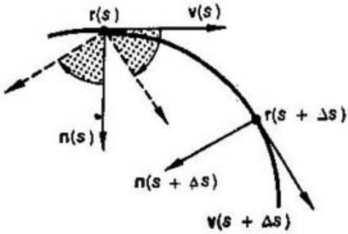

Figure 4.1

time, it follows from the Frenet formulas that $\frac{d}{ds}\omega(s)=\begin{pmatrix}0&k(s)\\-k(s)&0\end{pmatrix}\omega(s)$. Comparing these matrices, we see that $k(s)=d\varphi(s)/ds$. Thus, the curvature of a curve at a point s can be interpreted as the rate of the change in the angle $\varphi(s)$ at this point. In the case of a plane curve the function $k(s)$ completely defines the curve, provided $k\neq 0$ for all s. To be more exact, the following theorem is valid.

Theorem 3. *Given a smooth function $k(s)$ not vanishing for all s such that $a\leqslant s\leqslant b$. Then there exists a plane smooth curve $\mathbf{r}(s)$ uniquely defined to within a translation and an orthogonal transformation and such that for this curve $k(s)$ is the curvature and s is the natural parameter.*

Proof. Let us consider the system of differential equations $\begin{pmatrix}d\mathbf{v}(s)/ds\\d\mathbf{n}(s)/ds\end{pmatrix}=\begin{pmatrix}0&k(s)\\-k(s)&0\end{pmatrix}\begin{pmatrix}\mathbf{v}(s)\\\mathbf{n}(s)\end{pmatrix}$ where $k(s)$ is a given function. Since $k(s)\neq 0$, it follows from the existence and uniqueness theorem for differential equations that this system has a solution (unique for

fixed initial values) which can be continued smoothly to the entire interval $a < s < b$. Hence, we can consider the equation $\frac{dr(s)}{ds} = k(s)\,\mathbf{v}(s)$. According to the arguments just mentioned, it has a unique solution (for fixed initial values) which can be continued to the entire interval $a < s < b$. Since any two initial values on a plane are congruent relative to a translation and an orthogonal transformation, the unique solution is defined to within these transformations. The theorem is proved.

Special emphasis should be given to the condition $k(s) \neq 0$. Figure 4.2 shows two curves on a plane, which are described by the

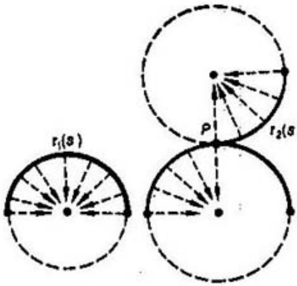

Figure 4.2

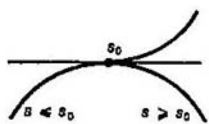

Figure 4.3

same smooth functions $k(s)$ but cannot, apparently, be transformed into each other by a translation or an orthogonal transformation. Both curves are composed of circular arcs. Distinct behaviour of these curves becomes especially clear if we consider the vector function of the normal $\mathbf{n}(s)$. The normal $\mathbf{n}_1(s)$ to the curve $\mathbf{r}_1(s)$ is always directed to the point O, the centre of the circle. The curve $\mathbf{r}_1(s)$ is smooth, the curve $\mathbf{r}_2(s)$ is not, for the second derivative of the radius vector $\mathbf{r}_2(s)$ suffers discontinuity at the point P. This is because the curvature $k(s)$ is non-zero and constant at all points of both curves. Yet, we can make two curves of the same curvature to be incongruent and smooth. The procedure is clear from the above example. Let us consider an arbitrary smooth curve whose curvature is a smooth function, vanishing at a point $s = s_0$ together with the derivatives of all orders. The existence of such a curve follows from Theorem 3 applied to a smooth function $k(s)$ which has a zero of infinite order at one of the ends of the interval $a \leqslant s \leqslant b$. Conjugating two such curves at their common end point, at which the curvatures vanish together with all the derivatives, we obtain two smooth incongruent curves with the same curvature functions (see Fig. 4.3).

4.1.2. THE THEORY OF SPATIAL CURVES. FRENET FORMULAS

Let us now consider a smooth curve $\mathbf{r}(t)$ in a Euclidean space $\mathbf{R}^n$ referred to Cartesian coordinates $x^1, \ldots, x^n$, i.e. $\mathbf{r}(t) = (x^1(t), \ldots, x^n(t))$. Just like in the planar case, with each point of $\mathbf{r}(s)$ we can uniquely associate a frame which is smoothly varying along the curve as the natural parameter s changes. Before constructing this frame, which we shall call "the Frenet frame", we shall prove, by analogy with the planar case, an auxiliary proposition about differentiation of a matrix-valued function.

Consider a smooth curve in a linear matrix space, i.e. a one-parameter family $A(t)$, where t varies on the interval $-a < t < a$, and $A(t)$ is an $n \times n$ matrix whose coefficients are smooth functions of t. Suppose all the matrices $A(t)$ (for all $-a < t < a$) are orthogonal and have $\det A(t) = +1$ and $A(0) = E$, where E is an identity matrix.

Lemma 3. *Let $X = \dot{A}(t)|_{t=0}$ denote the derivative of this one-parameter family of orthogonal matrices $A(t)$ at $t = 0$, i.e. X is the matrix composed of the functions $\left.\frac{da_{ij}(t)}{dt}\right|_{t=0}$, where $A(t) = (a_{ij}(t))$. Then X is a skew-symmetric matrix.*

Proof. Each orthogonal matrix $A(t)$ can be represented as a linear operator in an n-dimensional Euclidean space, the operator preserving the Euclidean scalar product. This means that for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ the identity $(A(t)\mathbf{x}, A(t)\mathbf{y}) = (\mathbf{x}, \mathbf{y})$ is valid. Since the left-hand side of the identity is a smooth function of t, we can calculate its derivative with respect to t at $t = 0$. We have $(\dot{A}(t)\mathbf{x}, A(t)\mathbf{y}) + (A(t)\mathbf{x}, \dot{A}(t)\mathbf{y})|_{t=0} = 0$, i.e. $(X\mathbf{x}, \mathbf{y}) + (\mathbf{x}, X\mathbf{y}) = 0$. Taking the basis vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ as $\mathbf{x}$ and $\mathbf{y}$, respectively ($\mathbf{e}_1, \ldots, \mathbf{e}_n$ is an orthogonal basis in $\mathbf{R}^n$), we obtain $(X\mathbf{e}_i, \mathbf{e}_j) + (\mathbf{e}_i, X\mathbf{e}_j) = 0$, which means that X is skew-symmetric in $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. The lemma is proved.

Remark. Since the set of orthogonal matrices is embedded as a subset in the linear space of all $n \times n$ matrices, $A(t)$ can be considered as a smooth curve in an n^2-dimensional linear space. From this point of view the matrix $X = \dot{A}(t)|_{t=0}$ is naturally identified with an ordinary velocity vector of a smooth curve $A(t)$ at $t = 0$.

A matrix-valued function $A(t)$ can be expanded at $t = 0$ in the powers of an infinitesimal increment

$$\Delta t \colon A(\Delta t) = E + \left.\frac{dA(t)}{dt}\right|_{t=0} \cdot \Delta t + \ldots.$$

The matrix X appears then as a "coefficient" of Δt. We now turn to constructing the Frenet frame.

Let $\mathbf{r}(s)$ be a smooth vector function which defines a smooth trajectory $\gamma(s)$ in $\mathbf{R}^n$. Suppose that for $a \leqslant s \leqslant b$ all vectors $\frac{d\mathbf{r}}{ds}$, $\frac{d^2\mathbf{r}}{ds^2}, \ldots, \frac{d^n\mathbf{r}}{ds^n}$ are linearly independent for each s. Then at every point of $\mathbf{r}(s)$ there exists a coordinate frame (not orthonormal!) which is composed of the vectors $\frac{d\mathbf{r}(s)}{ds}, \ldots, \frac{d^n\mathbf{r}(s)}{ds^n}$ and smoothly varies from point to point. Since the derivatives of the radius vector are linearly independent, all $\frac{d^k\mathbf{r}}{ds^k}$ are non-zero, $1 \leqslant k \leqslant n$.

Proposition 1. *Let* $\mathbf{r}(s)$ *be a smooth vector function in* $\mathbf{R}^n$ *and let the kth derivative* $\frac{d^k\mathbf{r}}{ds^k}$ *depend linearly on* $\frac{d\mathbf{r}}{ds}, \frac{d^2\mathbf{r}}{ds^2}, \ldots, \frac{d^{k-1}\mathbf{r}}{ds^{k-1}}$ *at each point of the interval* $a \leqslant s \leqslant b$. *Let also* $\frac{d^k\mathbf{r}}{ds^k} \neq 0$ *for* $a \leqslant s \leqslant b$ *and all* $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^{k-1}\mathbf{r}}{ds^{k-1}}$ *be linearly independent for* $a \leqslant s \leqslant b$. *Then the curve* $\mathbf{r}(s)$ *lies entirely in the* $(k-1)$*-dimensional plane spanned by the vectors* $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^{k-1}\mathbf{r}}{ds^{k-1}}$, *and this plane does not change its position in* $\mathbf{R}^n$ *as s varies from a to b.*

Proof. The hypothesis implies that there exist smooth functions $\lambda_i(s)$, $1 \leqslant i \leqslant k-1$, such that at each point s $\frac{d^k\mathbf{r}}{ds^k} = \sum_{i=1}^{k-1} \lambda_i(s) \frac{d^i\mathbf{r}}{ds^i}$.

Since the vectors $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^{k-1}\mathbf{r}}{ds^{k-1}}$ are linearly independent, they can be taken as a basis in the plane $\mathbf{R}^{n-1}$ spanned by these vectors. In order to prove that the curve $\mathbf{r}(s)$ always lies in the same k-dimensional plane, it is sufficient to demonstrate that the position of the plane $\mathbf{R}^{k-1}(s)$ in $\mathbf{R}^n$ remains unchanged as s varies. Since $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^{k-1}\mathbf{r}}{ds^{k-1}}$ are the basis vectors in $\mathbf{R}^{k-1}(s)$, it suffices to show that the derivative of the basis consists of the vectors which can be decomposed in the basis vectors. But this is obvious from the hypothesis, so the proposition is proved.

Thus, if the vectors $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^n\mathbf{r}}{ds}$ are linearly independent, the curve $\mathbf{r}(s)$ is not contained in any $(n-1)$-dimensional plane $\mathbf{R}^{n-1} \subset \mathbf{R}^n$ whose position remains unchanged when s varies. We now construct at each point s an orthonormal basis whose vectors are denoted by $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$. Put $\frac{d\mathbf{r}}{ds} \Big/ \left| \frac{d\mathbf{r}}{ds} \right| = \boldsymbol{\tau}_1$ and consider a two-dimensional plane spanned by $\boldsymbol{\tau}_1$ and $\frac{d^2\mathbf{r}}{ds^2}$ in which we choose a vector $\boldsymbol{\tau}_2$ orthogonal to $\boldsymbol{\tau}_1$. Since, by assumption, $\frac{d\mathbf{r}}{ds}$ and $\frac{d^2\mathbf{r}}{ds^2}$ are

linearly independent, τ_2 has a nonzero projection on $\frac{d^2\mathbf{r}}{ds^2}$. We shall choose the next vector τ_3 in a three-dimensional space spanned by $\frac{d^3\mathbf{r}}{ds^3}$, τ_2, τ_1, i.e. by the vectors $\frac{d^3\mathbf{r}}{\partial s^3}$, $\frac{d^2\mathbf{r}}{ds^2}$, $\frac{d\mathbf{r}}{ds}$, and take for τ_3 the vector orthogonal to the plane spanned by τ_2 and τ_1 $\left(\text{i.e. by } \frac{d^2\mathbf{r}}{ds^2} \text{ and } \frac{d\mathbf{r}}{ds}\right)$. Continuing the procedure, we obtain the orthonormal frame $\tau_1, \ldots, \tau_n$. Apparently, when s changes, the vectors of the frame $\tau(s)$ also vary smoothly from point to point. As in the planar case, it can be shown that the derivatives of the frame vectors $(\tau_1(s), \ldots, \tau_n(s))$ satisfy simple relations similar to the Frenet formulas for a plane curve.

Theorem 4. *Let a smooth curve* $\mathbf{r}(s)$ *in* $\mathbf{R}^n$ *be referred to the natural parameter* s *and let the vectors* $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^n\mathbf{r}}{ds^n}$ *be linearly independent at each point* $a \leqslant s \leqslant b$. *Then there exist functions* $k_2(s), \ldots, k_n(s)$ *such that the following relations (Frenet formulas) are satisfied:*

$$\begin{cases} \dfrac{d\tau_1}{ds} = k_2\tau_2, \\ \dfrac{d\tau_2}{ds} = k_3\tau_3 - k_2\tau_1, \\ \ldots\ldots\ldots\ldots\ldots \\ \dfrac{d\tau_{n-1}}{ds} = k_n\tau_n - k_{n-1}\tau_{n-2}, \\ \dfrac{d\tau_n}{ds} = \qquad\quad - k_n\tau_{n-1}. \end{cases}$$

i.e. $\frac{d\tau_i}{ds} = k_{i+1}\tau_{i+1} - k_i\tau_{i-1}$, *where* $k_1 = k_{n+1} = 0$.

Proof. Since the vector τ_i $(1 \leqslant i \leqslant n)$ is contained in the linear closure of $\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^i\mathbf{r}}{ds^i}$, its derivative $\frac{d\tau_i}{ds}$, is contained in the linear closure of the vectors

$$\frac{d\mathbf{r}}{ds}, \ldots, \frac{d^i\mathbf{r}}{ds^i}, \frac{d^{i+1}\mathbf{r}}{ds^{i+1}},$$

i.e. there exist functions $a_{i,1}(s), \ldots, a_{i,i+1}(s)$ such that

$$\frac{d\tau_i}{ds} = \sum_{j=1}^{i+1} a_{ij}(s)\,\tau_j(s).$$

Hence, the $n \times n$ matrix X, which represents the vectors $\frac{d\tau_1}{ds}, \ldots, \frac{d\tau_n}{ds}$ in terms of the coordinate frame

$$\tau(s) = (\tau_1, \ldots, \tau_n),$$

is of the form

$$X=\begin{pmatrix} a_{11} & a_{21} & & & & \\ a_{12} & a_{22} & a_{32} & & * & \\ & a_{23} & a_{33} & a_{43} & & \\ & & \ddots & & & \\ & & & a_{n-1,\, n-2} & & \\ & 0 & & a_{n-1,\, n-1} & a_{n,\, n-1} \\ & & & a_{n-1,\, n} & a_{nn} \end{pmatrix}.$$

The matrix X is known to have some additional properties. Indeed, as in the two-dimensional case, the translation of the frame $\tau(s)$ along the curve $\gamma(s)$ can be interpreted in terms of a family of orthogonal matrices such that $\tau(s)=A(s)\cdot\tau(0)$. This one-parameter family uniquely defines the evolution of $\tau(s)$ when s varies. Then, apparently, X will coincide with the derivative $\frac{dA(s)}{ds}$ at $s=0$. According to Lemma 3, the matrix X is skew-symmetric, i.e.

$$X=\begin{pmatrix} 0 & -a_{12} & & & & \\ a_{12} & 0 & & -a_{23} & 0 & \\ & a_{23} & 0 & -a_{34} & & \\ & & \ddots & & & \\ & & 0 & a_{n-2,\, n-1} & 0 & -a_{n-1,\, n} \\ & & & & a_{n-1,n} & 0 \end{pmatrix}.$$

Taking the functions $a_{i,\, i+1}(s)$ as $k_{i(s)}$, we complete the proof of the theorem.

We now turn to three dimensions, $n=3$. Here the Frenet formulas become

$$\frac{d\tau_1}{ds}=k_2\tau_2,\quad \frac{d\tau_2}{ds}=k_3\tau_3-k_2\tau_1,\quad \frac{d\tau_3}{ds}=-k_3\tau_2.$$

The vector τ_1 is a unit velocity vector of the curve $\mathbf{r}(s)$ and is usually denoted by $\mathbf{v}(s)$. The vector $\tau_2(s)$ coincides with the derivative of $\mathbf{v}$ with respect to s, since $\frac{d\mathbf{v}}{ds}\perp\mathbf{v}$ and $|\mathbf{v}(s)|=1=\text{const}$. Here we have used the fact that s is the natural parameter. The vector τ_3 is orthogonal both to $\mathbf{v}$ and $\mathbf{n}=\frac{1}{k}\frac{d\mathbf{v}}{ds}$, i.e. we may assume that it coincides with the vector product of $\mathbf{v}$ and $\mathbf{n}$. The vector $\mathbf{n}$ is called a *normal vector* to the curve $\mathbf{r}(s)$ and the vector $\mathbf{b}=[\mathbf{v},\mathbf{n}]$ a *binormal vector* to $\mathbf{r}(s)$. (Here $[\mathbf{v},\mathbf{n}]$ stands for the vector product of $\mathbf{v}$ and $\mathbf{n}$.)

In this notation the Frenet formulas take the form

$$\frac{d\mathbf{v}}{ds} = k\mathbf{n}, \quad \frac{d\mathbf{n}}{ds} = \varkappa\mathbf{b} - k\mathbf{v}, \quad \frac{d\mathbf{b}}{ds} = -\varkappa\mathbf{n},$$

where $k(s) = k_2(s)$ is called the *curvature* of a curve and $\varkappa(s) = k_3(s)$ is the *torsion* of a curve. (The vector $\mathbf{n}(s)$ is sometimes called a *principal normal* to $\mathbf{r}(s)$, rather than simply a normal.)

For the sake of convenience we shall always assume that $\boldsymbol{\tau}_2$ coincides with $\frac{d\mathbf{v}}{ds} \Big/ \left|\frac{d\mathbf{v}}{ds}\right|$; then $k(s)$ coincides with the magnitude of $\frac{d\mathbf{v}(s)}{ds}$ and is therefore a positive number (the corresponding derivatives of the radius vector are supposed to be non-zero). In the case of a plane curve the binormal vector is constant, i.e. it does not alter as a point moves along the curve; in particular, the torsion of the curve is zero. Hence, zero torsion may be considered as a characteristic of a plane curve in $\mathbf{R}^3$. We now discuss in greater detail the properties of a spatial curve with non-zero torsion. Consider a translation of the coordinate frame $(\mathbf{v}, \mathbf{n}, \mathbf{b})$ along a curve and project $\mathbf{b}(s)$, $\mathbf{n}(s)$, and $\mathbf{v}(s)$ onto the plane spanned by $\mathbf{b}(s_0)$ and $\mathbf{n}(s_0)$, where s_0 is a fixed value of the parameter s and the value of s is assumed to be infinitely close to s_0. In this case the velocity vector is projected as an infinitesimal vector, and we may assume, in the first approximation, that this vector vanishes. Then we have a motion of the vectors $\mathbf{b}(s)$, $\mathbf{n}(s)$ projected onto the plane spanned by $\mathbf{b}(s_0)$ and $\mathbf{n}(s_0)$. It follows from the Frenet formulas that this motion is described by the relations $\frac{d\mathbf{n}}{ds} = \varkappa\mathbf{b}$, $\frac{d\mathbf{b}}{ds} = -\varkappa\mathbf{n}$, i.e. the motion is determined by the skew-symmetric matrix $\begin{pmatrix} 0 & \varkappa \\ -\varkappa & 0 \end{pmatrix}$, which governs an infinitesimal rotation of the coordinate frame $\mathbf{b}$, $\mathbf{n}$. Hence, the vectors $\mathbf{b}$, $\mathbf{n}$ are rotated about the velocity vector of a curve, the speed of this rotation being uniquely determined by the torsion of the curve (incidentally, this is the reason for the term "torsion"). And if a curve was initially plane, it becomes "twisted in space". Thus, a spatial curve can (locally) be obtained from a plane curve if we move along the latter at a constant velocity and "twist" it at every moment by the torsion $\varkappa$ (Fig. 4.4).

Remark. As in the case of a plane curve, there exist invariants—curvature and torsion—which uniquely (to within the isometry of a three-dimensional space) define a smooth curve. Since we shall not use this assertion (which can be proved exactly in the same way as Theorem 3), the proof for the three-dimensional case is omitted.

Let us consider a smooth curve in $\mathbf{R}^n$ and let the velocity vector of this curve be non-zero at each point.

Proposition 2. *A smooth curve with non-zero velocity vector is a smooth one-dimensional manifold smoothly embedded in* $\mathbf{R}^n$.

Proof. That the curve is a smooth manifold is a straightforward consequence of the definition of a smooth manifold. Thus, it remains to verify that the curve is a smooth submanifold in $\mathbf{R}^n$. To this end

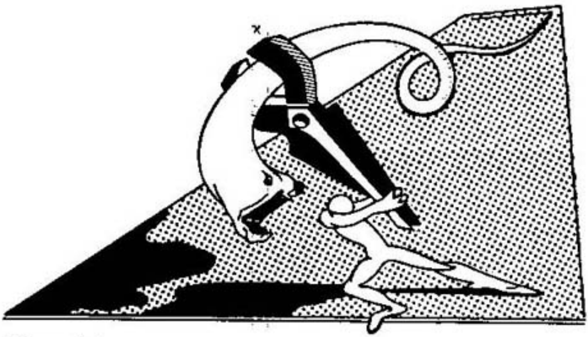

Figure 4.4

we have to consider the differential of embedding i, i.e. the linear mapping di. Since this mapping is completely defined by the velocity vector, the Jacobi matrix has a maximal rank at each point of the curve, so that the curve is a submanifold.

Remark. The plane curve shown in Fig. 4.5 is a smooth manifold, but it is not a smooth submanifold in a plane. Obviously, this curve

Figure 4.5

can be defined by a radius vector with smooth components $x(t)$ and $y(t)$, but at the vertices of the triangle the velocity vector must vanish in order that the curve may sharply change its direction.

Apparently, any one-dimensional smooth, closed (without boundary) manifold is diffeomorphic either to a straight line (noncompact manifold) or to a circle (compact manifold). Thus, the

class of one-dimensional manifolds consists only of two manifolds considered to within a diffeomorphism. They are not diffeomorphic, for a straight line is non-compact while a circle is compact.

Problems

1. Prove that if the curvature of a plane curve is identically zero, this curve is a straight line.
2. Prove that a plane curve in a three-dimensional space is characterized by the condition $\varkappa \equiv 0$.
3. Prove that a straight line in a three-dimensional space is characterized by the conditions: $k \equiv 0$, $\varkappa \equiv 0$.
4. Describe the class of those curves for which $k = \text{const}$ and $\varkappa = \text{const}$.
5. Prove that the trajectory of the motion of a particle in a central field is a plane curve.

4.2. SURFACES. FIRST AND SECOND FUNDAMENTAL FORMS

4.2.1. THE FIRST FUNDAMENTAL FORM

Let us consider a Euclidean space $\mathbf{R}^n$ and let V^{n-1} be a smooth manifold of dimension $n - 1$ (or of codimension one) embedded in $\mathbf{R}^n$. In this section we shall mainly concentrate on the local properties of this hypersurface, leaving aside its global structure. We may assume therefore that we deal with a smooth embedding of a disk D^{n-1} in $\mathbf{R}^n$. We have already discussed different ways of defining a manifold, including a hypersurface. For the sake of convenience, we shall use the parametric definition of V^{n-1}, that is, we shall assume V^{n-1} (or embedded disk D^{n-1}) to be defined by a smooth radius vector $\mathbf{r} = \mathbf{r}(u^1, \ldots, u^{n-1}) \subset \mathbf{R}^n$ where the parameters (coordinates) $u^1, \ldots, u^{n-1}$ run some disk in a Euclidean parameter space $\mathbf{R}^{n-1}(u^1, \ldots, u^{n-1})$. Since the radius vector is assumed to define a smooth submanifold, the vectors $\frac{\partial \mathbf{r}}{\partial u^1}, \ldots, \frac{\partial \mathbf{r}}{\partial u^{n-1}}$ are linearly independent at each point of the domain of their definition. Recall that these vectors are tangent to the corresponding coordinate lines through a point P on V^{n-1}. As was shown above, a smooth embedding of V^{n-1} in $\mathbf{R}^n$ induces on V^{n-1} a Riemannian metric. We now remind the construction. Let $x^1, \ldots, x^n$ be Cartesian coordinates in $\mathbf{R}^n$, then the radius vector $\mathbf{r}$ is given by the set of smooth functions $x^i(u^1, \ldots, u^{n-1})$, $1 \leqslant i \leqslant n$. Let $ds^2 = \sum_{i=1}^{n} (dx^i)^2$ be the Euclidean metric in $\mathbf{R}^n$; then we have the follow-

ing form:

$$ds^2\Big|_{V^{n-1}} = \sum_{i=1}^{n} (dx^i(u^1, \ldots, u^{n-1}))^2 = \sum_{i=1}^{n} \sum_{k=1}^{n-1} \left(\frac{\partial x^i}{\partial u^k} du^k\right)^2$$

$$= \sum_{i=1}^{n} \sum_{k,p=1}^{n-1} \frac{\partial x^i}{\partial u^k} \cdot \frac{\partial x^i}{\partial u^p} du^k du^p = \sum_{k,p=1}^{n-1} g_{kp}(u) du^k du^p,$$

$$g_{kp} = \left\langle \frac{\partial \mathbf{r}}{\partial u^k}, \frac{\partial \mathbf{r}}{\partial u^p} \right\rangle,$$

where $\langle , \rangle$ stands for the scalar product in $\mathbf{R}^n$.

Definition. *The first fundamental form of a hypersurface V^{n-1} in $\mathbf{R}^n$ is the form $ds^2\Big|_{V^{n-1}} = \sum_{k,p} g_{kp} du^k du^p$, where the functions $g_{kp}(u^1, \ldots, u^{n-1})$ have been defined above.*

These functions depend on the radius vector of the hypersurface and vary, in general, on the change of the radius vector, i.e. upon the deformation of the hypersurface. The first fundamental form is defined on the vectors tangent to V^{n-1}; to be more exact, if $\mathbf{a}, \mathbf{b} \in T_P V^{n-1}$ are two arbitrary tangent vectors, there is defined the scalar product $\langle \mathbf{a}, \mathbf{b} \rangle_{ds^2(V^{n-1})} = g_{kp} a^k b^p = \sum_{k,p=1}^{n-1} g_{kp} a^k b^p$. Note that to simplify the notation we omit the symbol $\sum$ if summation is performed over the same upper and lower indices. It follows from the geometric meaning of g_{kp} that the scalar product $\langle \mathbf{a}, \mathbf{b} \rangle_{ds^2(V^{n-1})}$ coincides with the ordinary scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$ considered as vectors of the ambient space $\mathbf{R}^n$. The metric tensor matrix $\mathfrak{G}$ composed of the functions $g_{kp}(u^1, \ldots, u^{n-1})$ is of the form

$$\mathfrak{G} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1,n-1} \\ \cdot & \cdot & \cdots & \cdot \\ g_{n-1,1} & g_{n-1,2} & \cdots & g_{n-1,n-1} \end{pmatrix}.$$

It should be noted that the Riemannian metric has appeared earlier as a convenient tool for calculating the length of a smooth curve on a surface. From this point of view, the length of an arc $\gamma(t)$ in V^{n-1} is given by $l_a^b(\gamma(t)) = \int_a^b \sqrt{\langle \dot{\gamma}, \dot{\gamma} \rangle}\, dt$. In terms of $g_{kp}(u)$ this expression takes the form

$$l_a^b \gamma(t) = \int_a^b \sqrt{\sum_{k,p=1}^{n-1} g_{kp}(u(t)) \frac{du^k(t)}{dt} \frac{du^p(t)}{dt}}\, dt.$$

This formula is more preferable in that it contains the functions $g_{kp}(u)$ which are independent of the choice of a curve on V^{n-1}, and only depend on the surface V^{n-1} and its parametrization. Furthermore, these functions do not vary under isometries of a surface; for example, when a plane is bent into a cylinder.

We now study how the first fundamental form is expressed for various definitions of a hypersurface. Let V^{n-1} be given as a graph $x^n = f(x^1, \ldots, x^{n-1})$. We have

$$\begin{aligned} ds^2\Big|_{V^{n-1}} &= \sum_{i=1}^{n-1} (dx^i)^2 + (dx^n(x^1, \ldots, x^{n-1}))^2 \\ &= \sum_{i=1}^{n-1} (dx^i)^2 + \sum_{k,\,p=1}^{n-1} \frac{\partial f}{\partial x^k} \frac{\partial f}{\partial x^p} dx^k\, dx^p \\ &= \sum_{k,\,p=1}^{n-1} \left(\delta_{kp} + \frac{\partial f}{\partial x^k} \frac{\partial f}{\partial x^p}\right) \cdot dx^k\, dx^p \\ &= g_{kp}(x^1, \ldots, x^{n-1})\, dx^k\, dx^p. \end{aligned}$$

(The derivative $\frac{\partial f}{\partial x^\alpha}$ will sometimes be denoted f_{x^α}.) Let now V^{n-1} be given as an implicit function, i.e. in the form $F(x^1, \ldots, x^n) = 0$, where $\frac{\partial F}{\partial x^n} \neq 0$. Then, it follows from the implicit function theorem that the equation $F(x^1, \ldots, x^n) = 0$ has (locally) the solution $x^n = f(x^1, \ldots, x^{n-1})$, with $\partial f/\partial x^i = -(\partial F/\partial x^i)/(\partial F/\partial x^n)$. Substituting $(\partial F/\partial x^\alpha)/(\partial F/\partial x^n)$ for f_{x^α}, we obtain $\mathfrak{G} = (g_{kp})$, where

$$g_{kp} = \left[\left(\frac{\partial F}{\partial x^k} \cdot \frac{\partial F}{\partial x^p}\right) \Big/ \left(\frac{\partial F}{\partial x^n}\right)^2\right] + \delta_{kp}.$$

Let us consider a particular case: embedding of a two-dimensional surface in three-dimensional Euclidean space. Let V^2 be given parametrically: $\mathbf{r} = \mathbf{r}(u, v)$. Then the first fundamental form is usually written as $ds^2(V^2) = Edu^2 + 2Fdudv + Gdv^2$, where $E = \langle \mathbf{r}_u, \mathbf{r}_u \rangle$, $F = \langle \mathbf{r}_u, \mathbf{r}_v \rangle$, and $G = \langle \mathbf{r}_v, \mathbf{r}_v \rangle$ are the coefficients of the form in terms of the components of the radius vector $\mathbf{r}$:

$E = x_u^2 + y_u^2 + z_u^2$, $F = x_u x_v + y_u y_v + z_u z_v$, $G = x_v^2 + y_v^2 + z_v^2$. The metric $ds^2(V^2)$ is called conformal-Euclidean if $E = G$ and $F = 0$.

Here is an example of the first fundamental form for a surface of rotation embedded in $\mathbf{R}^3$. Let the cylindrical coordinates (r, φ, z) be valid in $\mathbf{R}^3$ and let a two-dimensional surface be defined parametrically, $(\varphi = u,\ z = v,\ r = r(v))$. Calculation yields

$$ds^2(V^2) = dv^2 + r^2(v)\, du^2 + (dr(v))^2 = (1 + (r_v')^2)\, dv^2 + r^2\, du^2.$$

Here $F(u, v) = 0$, $E(u, v) = r^2(v)$, $G(u, v)' = 1 + (r'_v)^2$. That $F = 0$ means that the coordinate lines $v = v_0 = \text{const}$ and $u = u_0 = \text{const}$ are orthogonal at each point (Fig. 4.6).

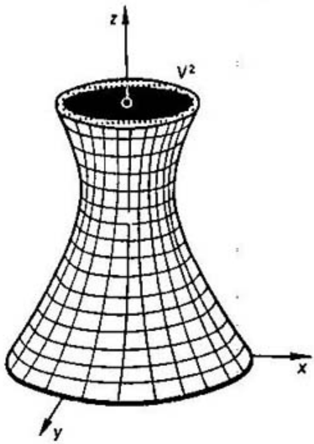

Figure 4.6

Lemma 1. *Let $V^{n-1} \subset \mathbf{R}^n$ be a smooth submanifold and let $\mathfrak{G}$ be the first fundamental form. Then $\mathfrak{G}$ is non-singular.*

Proof. By the definition of the radius vector $\mathbf{r} = \mathbf{r}(u^1, \ldots, u^{n-1})$, all vectors $\mathbf{r}_{u^k}$, $1 \leqslant k \leqslant n - 1$, are linearly independent at the points P belonging to V^{n-1}. Since the matrix $\mathfrak{G}$ is composed of the scalar products of $\mathbf{r}_{u^k}$ and $\mathbf{r}_{u^p}$, i.e. $\mathfrak{G} = (\langle \mathbf{r}_{u^k}, \mathbf{r}_{u^p}\rangle)$, $\mathfrak{G}$ is non-singular. The lemma is proved.

This fact is easily seen from geometric considerations. Indeed, if the symmetric matrix $\mathfrak{G}$ were singular, it would have at least one zero eigenvalue, and then the initial Euclidean metric would also have a zero eigenvalue, which contradicts the definition of this metric.

4.2.2. THE SECOND FUNDAMENTAL FORM

Let V^{n-1} be a hypersurface in $\mathbf{R}^n$ given by the radius vector $\mathbf{r} = \mathbf{r}(u^1, \ldots, u^{n-1})$ and let $\mathbf{n} = \mathbf{n}(P)$ be a unit vector orthogonal to V^{n-1} at a point P. We introduce the fundamental form $Q(\mathbf{a}, \mathbf{a})$ by defining its values for an arbitrary vector $\mathbf{a} \in T_P(V^{n-1})$.

To this end, we consider an arbitrary smooth curve $\gamma(t)$ on V^{n-1} through the point P such that $\gamma(0) = P$, $\dot{\gamma}(0) = \mathbf{a}$. Such a curve always exists, though it is not defined uniquely (Fig. 4.7). Since the radius vector $\mathbf{r}$ is a function of t along the curve $\gamma(t)$, then $\mathbf{a} = \frac{d}{dt}\mathbf{r}(u(t))|_{t=0}$. We consider a vector function $\dot{\mathbf{r}} = \frac{d}{dt}\mathbf{r}(u(t))$ and its derivative with respect to t, i.e. $\ddot{\mathbf{r}} = \frac{d^2}{dt^2}\mathbf{r}(u(t))$. Let $\ddot{\mathbf{r}}_{\mathbf{a}}$ stand for

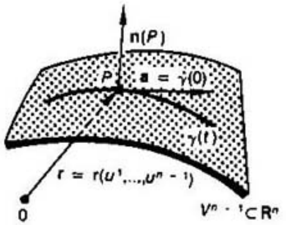

Figure 4.7

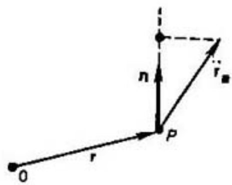

Figure 4.8

the value of $\ddot{\mathbf{r}}$ at $t = 0$. This is the second derivative of the radius vector $\mathbf{r}$ along the direction of the vector $\mathbf{a}$.

Definition. Put $Q(\mathbf{a}, \mathbf{a}) = \langle \ddot{\mathbf{r}}_{\mathbf{a}}, \mathbf{n}\rangle$.

The number thus defined is the projection of $\ddot{\mathbf{r}}_{\mathbf{a}}$ onto the normal vector $\mathbf{n}$ at point P (see Fig. 4.8). Calculate $Q(\mathbf{a}, \mathbf{a})$ explicitly in terms of the coordinates of $\mathbf{r}$. We have

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}_{u^k}\frac{du^k(t)}{dt}, \quad \frac{d^2\mathbf{r}}{dt^2} = \mathbf{r}_{u^k u^p}\frac{du^k}{dt}\cdot\frac{du^p}{dt} + \mathbf{r}_{u^k}\frac{d^2u^k}{dt^2},$$

$$\left\langle \frac{d^2}{dt^2}\mathbf{r}(u(t)), \mathbf{n}\right\rangle = \left\langle \mathbf{r}_{u^k u^p}\frac{du^k}{dt}\frac{du^p}{dt}, \mathbf{n}\right\rangle + 0,$$

$$\left\langle \mathbf{r}_{u^k}\frac{d^2u^k}{dt^2}, \mathbf{n}\right\rangle = \frac{d^2u^k}{dt^2}\langle \mathbf{r}_{u^k}, \mathbf{n}\rangle = 0,$$

$$\mathbf{r}_{u^k} \in T_P(V^{n-1}), \quad \mathbf{n} \perp T_P(V^{n-1}),$$

$$\left\langle \frac{d^2\mathbf{r}}{dt^2}\Big|_{t=0}, \mathbf{n}\right\rangle = \langle \mathbf{r}_{u^k u^p}|_{t=0}, \mathbf{n}\rangle \frac{du^k}{dt}\Big|_{t=0}\frac{du^p}{dt}\Big|_{t=0} = Q(\mathbf{a}, \mathbf{a}).$$

It should be recalled that the vector $\mathbf{a}$ has the coordinates $\left(\frac{du^1(0)}{dt}, \ldots, \frac{du^{n-1}(0)}{dt}\right)$, that is, we finally obtain

$$Q(\mathbf{a}, \mathbf{a}) = \langle \mathbf{n}, \mathbf{r}_{u^k u^p}|_{t=0}\rangle a^k a^p.$$

This form uniquely defines a bilinear form $Q(\mathbf{a}, \mathbf{b})$ whose value on a pair of arbitrary vectors $\mathbf{a}, \mathbf{b} \in T_P V^{n-1}$ is $Q(\mathbf{a}, \mathbf{b}) = q_{kp}(P)\, a^k b^p$ where

$$q_{kp}(P) = \langle \mathbf{r}_{u^k u^p}\big|_{t=0}, \ \mathbf{n}\rangle .$$

Lemma 2. *The expression* $Q(\mathbf{a}, \mathbf{b}) = q_{kp} a^k b^p$, *where* $\mathbf{a}, \mathbf{b} \in T_P V^{n-1}$, *defines a bilinear form.*

Proof. We have

$$q_{k'p'} = \langle \mathbf{r}_{u^{k'} u^{p'}}\big|_{t=0}, \ \mathbf{n}\rangle = \frac{\partial u^k}{\partial u^{k'}} \frac{\partial u^p}{\partial u^{p'}} \left\langle \frac{\partial^2 \mathbf{r}(0)}{\partial u^k \partial u^p}, \ \mathbf{n} \right\rangle = \frac{\partial u^k}{\partial u^{k'}} \frac{\partial u^p}{\partial u^{p'}} q_{kp},$$

i.e. the functions q_{kp} are transformed under coordinate transformation as the coefficients of a bilinear form, which is what was required.

Definition. The bilinear form $Q(\mathbf{a}, \mathbf{b})$ is called the *second fundamental form of the surface* $V^{n-1} \subset \mathbf{R}^n$.

Obviously, the form Q depends on the way V^{n-1} is embedded in $\mathbf{R}^n$, that is, Q will, in general, vary under smooth deformation of V^{n-1}.

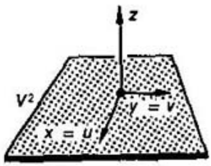

Figure 4.9

This form is not invariant under isometries of V^{n-1} in $\mathbf{R}^n$. For instance, on folding of V^{n-1}, i.e. under such a smooth deformation of $\mathbf{R}^n$ that the first fundamental form does not vary, the form Q will, in general, change. Let V^2 be a two-dimensional plane in $\mathbf{R}^3$, then the radius vector $\mathbf{r}(u, v)$ can be considered as a linear function of u and v. Hence, the first fundamental form is the Euclidean metric on a plane $du^2 + dv^2$ (Fig. 4.9). We now consider an isometric transformation of V^2, folding of a plane $\mathbf{R}^2$ into a cylinder with the axis parallel to the x-axis (Fig. 4.10). Apparently, the second fundamental form of a cylinder is not zero, since the number $\langle \mathbf{r}_{vv}, \mathbf{n}\rangle$ is not zero. At the same time, the second fundamental form of a plane $\mathbf{R}^2$ is identically zero, hence folding just mentioned (i.e. isometry) results in the change of this form.

Let us consider a fixed submanifold $V^{n-1} \subset \mathbf{R}^n$ and let $P \in V^{n-1}$. A pair of forms, $\mathfrak{G}$ and Q (the first and second fundamental forms), is defined at each point P. This pair is characterized by a set of numerical invariants which permit the study of V^{n-1} irrespective of

the coordinate system. Denote the matrices of the corresponding forms by $\mathfrak{G}$ and Q and consider the polynomial in λ, $\det(Q - \lambda\mathfrak{G})$. Since $\mathfrak{G}$ is non-singular (see Lemma 1), there exists a matrix $\mathfrak{G}^{-1}$ inverse to $\mathfrak{G}$, so that the equation $\det(\mathfrak{G}^{-1}Q - \lambda E) = 0$ is equivalent to $\det(Q - \lambda\mathfrak{G}) = 0$. Let $\lambda_1, \ldots, \lambda_{n-1}$ stand for the eigenvalues of the matrix $\mathfrak{G}^{-1}Q$, i.e. for the roots of the equation

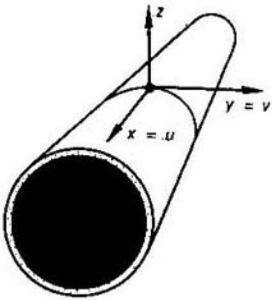

Figure 4.10

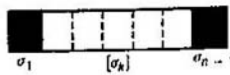

Figure 4.11

$\det(Q - \lambda\mathfrak{G}) = 0$. We shall prove in the sequel that they are all real. Write the characteristic polynomial $F(\lambda)$ as $\sum_{k=0}^{n-1} \sigma_k \lambda^k$, where σ_k are symmetric functions of $\lambda_1, \ldots, \lambda_{n-1}$.

Lemma 3. *The functions $\sigma_k(\lambda_1, \ldots, \lambda_{n-1})$ are invariants of the forms $\mathfrak{G}$ and Q, i.e. they are preserved under an arbitrary regular coordinate transformation in a neighbourhood of point $P \in V^{n-1}$.*

Proof. The regular coordinate transformation $x \to x'$ in the neighbourhood of point P in the tangent space $T_P V^{n-1}$ induces a linear regular transformation through the Jacobi matrix: $J: T_P \to T_P$, so that the matrices $\mathfrak{G}$ and Q are mapped as follows: $\mathfrak{G} \to J^T \mathfrak{G} J = \mathfrak{G}'$ and $Q \to J^T Q J = Q'$. Hence,

$$\det((\mathfrak{G}')^{-1}Q' - \lambda E) = \det[J^{-1} \cdot (\mathfrak{G}^{-1}Q - \lambda E) J] = \det(\mathfrak{G}^{-1}Q - \lambda E),$$

which is what was required.

Special attention will be given to the invariants

$$(1)\ \ \sigma_1 = \sum_{k=1}^{n-1} \lambda_k = \operatorname{spur}(\mathfrak{G}^{-1}Q), \qquad (2)\ \ \sigma_{n-1} = \prod_{k=1}^{n-1} \lambda_k = \det(\mathfrak{G}^{-1}Q).$$

The remaining σ_k, $2 \leqslant k \leqslant n-2$, describe more subtle properties

of V^{n-1}, which are now beyond our scope. Since $\mathfrak{G}^{-1}Q \neq 0$ for $Q \neq 0$, at least one of σ_k is non-zero. The invariants σ_1 and σ_{n-1} are "extreme" invariants (Fig. 4.11).

Definition. The function $H(P) = \sigma_1(P) = \sigma_1(\lambda_1, \ldots, \lambda_{n-1})$ is called the *mean curvature of a surface $V^{n-1} \subset \mathbf{R}^n$ at a point $P \in V^{n-1}$*. The function $K(P) = \sigma_{n-1}(P) = \sigma_{n-1}(\lambda_1, \ldots, \lambda_{n-1})$ is called the *Gaussian curvature of a surface $V^{n-1} \subset \mathbf{R}^n$* at a point P.

If $n = 3$ and $V^2 \subset \mathbf{R}^3$, then $H(P) = \lambda_1 + \lambda_2$; $K(P) = \lambda_1 \cdot \lambda_2$.

Theorem 1. *All eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$ of the pair of forms $\mathfrak{G}$, Q are real. If all $\lambda_1, \ldots, \lambda_{n-1}$ are pairwise distinct, all eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ of the matrix $\mathfrak{G}^{-1}Q$ are mutually orthogonal relative to both the ambient Euclidean metric in $\mathbf{R}^n$ and the Riemannian metric induced on V^{n-1} by the embedding $V^{n-1} \to \mathbf{R}^n$.*

Proof. According to the well-known algebraic theorem, the eigenvalues of a symmetric matrix are real and all its eigenvectors are mutually orthogonal for distinct eigenvalues. This theorem cannot be applied directly to our case, for the matrix $\mathfrak{G}^{-1}Q$ is not, in general, symmetric. It would be symmetric if $\mathfrak{G}$ and Q were commutative. Since the form $\mathfrak{G}(P)$ is symmetric at each P, there exists, in a certain neighbourhood of P, a regular coordinate transformation $x \to x'$ such that $\mathfrak{G}(P)$ can be diagonalized at a single point P. That is why a linear coordinate transformation appears to be sufficient. After $\mathfrak{G}$ is diagonalized, it can be reduced to an identity matrix through the extension along the principal axes of the form. Let A be a linear operator $A: T_PV^{n-1}$ which turns $\mathfrak{G}$ into an identity matrix, then $\mathfrak{G} = A^TEA = A^TA$, where the matrix E defines an orthogonal basis $\varphi_1, \ldots, \varphi_{n-1}$ in T_PV^{n-1}. Thus, we obtain $\det(Q - \lambda\mathfrak{G}) = \det[A^T((A^T)^{-1}Q(A^{-1}) - \lambda E)A]$. Consider the form $\widetilde{Q} = B^TQB$, where $B = A^{-1}$. The initial equation $\det(Q - \lambda\mathfrak{G}) = 0$ is written in the basis $\varphi_1, \ldots, \varphi_{n-1}$ as $\det(\widetilde{Q} - \lambda E) = 0$, since $\det A \neq 0$. We also have $\widetilde{Q}^T = \widetilde{Q}$, since $Q^T = Q$. Hence, the form $\widetilde{Q}$ and matrix $\mathfrak{G}^{-1}Q$ have the same eigenvalues and eigenvectors. Since $\widetilde{Q}$ is symmetric, all its eigenvalues (i.e. $\lambda_1, \ldots, \lambda_{n-1}$) are real, and if they are pairwise distinct all eigenvectors of Q are mutually orthogonal. This follows from the familiar algebraic theorem (the orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ need not necessarily coincide with $\varphi_1, \ldots, \varphi_{n-1}$). Theorem 1 is proved.

The proof implies that if all the eigenvalues are distinct, the orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ in the plane T_PV^{n-1} is defined uniquely. If some of $\lambda_1, \ldots, \lambda_{n-1}$ coincide, the relation will hold: $\mathbf{e}_i \perp \mathbf{e}_j$ for $\lambda_i \neq \lambda_j$. If some eigenvalue is of multiplicity k, there exists a k-dimensional invariant subspace such that its vectors are multiplied by the same number λ; in this subspace we can also choose (not uniquely!) a k-dimensional orthogonal basis.

Definition. The vectors $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ (defined uniquely if $\lambda_i \neq \lambda_j$ for $i \neq j$) are called *principal directions* (or *principal axes*) *of a hypersurface* V^{n-1} at a point P.

Thus, with each point of the hypersurface V^{n-1} we have associated a unique (if $\lambda_i \neq \lambda_j$ for $i \neq j$) orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ smoothly dependent on P. The vectors $\{\mathbf{e}_i\}$ are orthogonal both in the

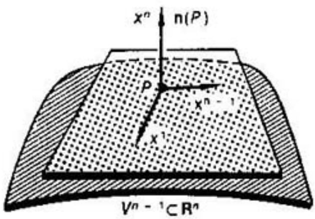

Figure 4.12

ambient Euclidean metric and in the induced metric on $V^{n-1} \subset \mathbf{R}^n$.

Since $\mathfrak{G}$ becomes an identity matrix in the orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$, the numbers $\lambda_1, \ldots, \lambda_{n-1}$ coincide with the eigenvalues of the form Q written in $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$.

Let us consider a particular case. Let $V^{n-1} \subset \mathbf{R}^n$ be given as the graph $x^n = f(x^1, \ldots, x^{n-1})$ and let at a point $P \in V^{n-1}$ the plane $T_P V^{n-1}$ coincide with the plane of the variables $x^1, \ldots, x^{n-1}$ (Fig. 4.12). Then the normal $\mathbf{n}(P)$ to V^{n-1} at the point P has the coordinates $(0, \ldots, 0, 1)$ and the radius vector $\mathbf{r}$ describing V^{n-1} is of the form $\mathbf{r} = \mathbf{r}(x^1, \ldots, x^{n-1}) = (x^1, \ldots, x^{n-1}; f(x^1, \ldots, x^{n-1}))$. Since the hyperplane $(x^1, \ldots, x^{n-1}) = T_P V^{n-1}$ is tangent to V^{n-1} at P, the relation $\partial f/\partial x^i|_P = 0$, $1 \leqslant i \leqslant n-1$, is satisfied, whence $\mathfrak{G}(P) = E$, since $g_{ij} = f_{x^i} f_{x^j} + \delta_{ij}$. Consider the matrix $Q = (q_{ij})$, where

$$q_{ij} = \langle \mathbf{r}_{x^i x^j}, \mathbf{n} \rangle = \frac{\partial^2 f(P)}{\partial x^i \partial x^j}.$$

Thus, $Q = (f_{x^i x^j}(P))$ coincides with the Hessian of f at the point P. The mean curvature $H(P)$ is

$$H(P) = \operatorname{spur}(\mathfrak{G}^{-1} Q) = \operatorname{spur} Q = \sum_{k=1}^{n-1} f_{x^k x^k};$$

and the Gaussian curvature is

$$K(P) = \det(\mathfrak{G}^{-1} Q) = \det Q = \det(f_{x^i x^j}(P)).$$

For a two-dimensional surface ($n=3$) we obtain

$$z=f(x,\ y),$$
$$H(P)=\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=\Delta f,$$

where $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$ is the Laplacian on the plane $\mathbf{R}^2$;

$$K(P)=\det\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy}\end{pmatrix}=f_{xx}f_{yy}-f_{xy}^2.$$

4.2.3. AN ELEMENTARY THEORY OF SMOOTH CURVES ON A HYPERSURFACE

Let us consider an arbitrary point $P\in V^{n-1}$ and let $\mathbf{n}(P)$ be a normal to the hypersurface V^{n-1} in $\mathbf{R}^n$. Consider also an arbitrary two-dimensional plane $\mathbf{R}^2$ through the point P which intersects V^{n-1} along a smooth curve $\gamma(t)=\mathbf{R}^2\cap V^{n-1}$ (as previously, we are only interested in a small neighbourhood of P).

Definition. The smooth curve $\gamma(t)=\mathbf{R}^2\cap V^{n-1}$ is called a *plane section of a hypersurface* $V^{n-1}\subset\mathbf{R}^n$.

Infinitely many plane sections pass through the point P. At the same time, not every smooth curve $\gamma\subset V^{n-1}$ is a plane section of V^{n-1}. For the curve $\gamma\subset V^{n-1}$ not to be a plane section it is sufficient that γ have a non-zero torsion. Let γ be a plane section of V^{n-1} (the plane $\mathbf{R}^2$ containing γ is fixed) and let the point O, the origin of the radius vector $\mathbf{r}=\mathbf{r}(u^1,\ \ldots,\ u^{n-1})$ describing V^{n-1}, belong to $\mathbf{R}^2$. Let $\mathbf{m}(P)$ be a normal vector to the plane curve $\gamma=\mathbf{R}^2\cap V^{n-1}$ contained in $\mathbf{R}^2$ (Fig. 4.13). Since we consider an arbitrary section, the normals $\mathbf{n}$ and $\mathbf{m}$ may not in general coincide.

Introduce on $\gamma=\mathbf{R}^2\cap V^{n-1}$ the natural parameter s: $\gamma=\gamma(s)$ (s is the arc length). Then the curve $\gamma(s)$ in $\mathbf{R}^2$ is given by the radius vector $\mathbf{r}=\mathbf{r}(s)$, where $\mathbf{r}(s)=\mathbf{r}(u^1(s),\ \ldots,\ u^{n-1}(s))$. According to the Frenet formulas for plane curves, we have $k(s)=\left|\frac{d^2\mathbf{r}(s)}{d^2s}\right|$, where $k(s)$ is the curvature of $\gamma(s)$ at the point P. Recall also that $\frac{d^2\mathbf{r}(s)}{ds^2}=\mathbf{m}k(s)$, where $\mathbf{m}=\mathbf{m}(P)$, $P\in\gamma(s)$. On the other hand, if $\mathbf{a}=\frac{d}{ds}\mathbf{r}(s)$ is the velocity vector of the curve $\gamma(s)$ at P, we have (by the definition of the second fundamental form Q)

$$Q(\mathbf{a},\ \mathbf{a})=\left\langle\frac{d^2\mathbf{r}}{ds^2},\ \mathbf{n}\right\rangle=\langle k\mathbf{m},\ \mathbf{n}\rangle=k\langle\mathbf{m},\ \mathbf{n}\rangle=k\cos\theta,$$

where θ is the angle between $\mathbf{m}$ and $\mathbf{n}$ at the point P (see Fig. 4.13). On the other hand,

$$Q(\mathbf{a},\ \mathbf{a}) = Q\left(\frac{d\mathbf{r}}{ds},\ \frac{d\mathbf{r}}{ds}\right)$$
$$= Q\left(\frac{d\mathbf{r}}{dt},\ \frac{d\mathbf{r}}{dt}\right)\left(\frac{dt}{ds}\right)^2 = \left(\frac{ds}{dt}\right)^{-2}\cdot Q\left(\frac{d\mathbf{r}}{dt},\ \frac{d\mathbf{r}}{dt}\right),$$

where t is an arbitrary smooth parameter along the curve $\gamma = \mathbf{R}^2 \cap V^{n-1}$. Here $\frac{d\mathbf{r}}{dt} = \alpha$ is an arbitrary tangent vector to the curve γ at the point P. If $\alpha = (\alpha^1, \ldots, \alpha^{n-1})$, then

$$k\cos\theta = \frac{Q(\alpha,\ \alpha)}{g_{ij}\alpha^i\alpha^j} = \frac{q_{ij}\alpha^i\alpha^j}{g_{ij}\alpha^i\alpha^j}.$$

Since the curve γ, which is a plane section, can be drawn along any tangent vector $\alpha \in T_PV^{n-1}$, the following theorem is proved.

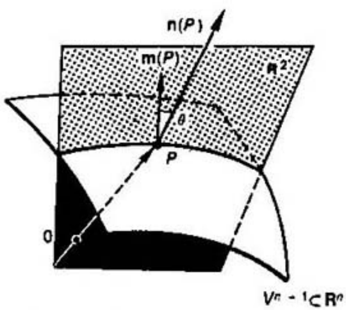

Figure 4.13

Theorem 2. *For any tangent vector $\alpha \in T_PV^{n-1}$ and any plane section γ (such that $\dot{\gamma} = \alpha$) the ratio of the second to the first fundamental form is $k\cdot\cos\theta$, i.e.*

$$\frac{Q(\alpha,\ \alpha)}{\mathfrak{G}(\alpha,\ \alpha)} = \frac{q_{ij}\alpha^i\alpha^j}{g_{ij}\alpha^i\alpha^j} = k\cos\theta.$$

The curvature k is called the *curvature of a plane section*.

Among plane sections, of special interest is the class of normal sections.

Definition. The plane section $\gamma = \mathbf{R}^2 \cap V^{n-1}$ at a point P is called *normal* if $\mathbf{n}(P) \in \mathbf{R}^2$, i.e. $\theta = 0,\ \pi$.

Thus, every normal section γ at a point $P \in V^{n-1}$ is uniquely defined by the tangent vector $\alpha \in T_PV^{n-1}$, i.e. the two-dimensional

plane $\mathbf{R}^2$ defining this section is spanned by two vectors, the normal $\mathbf{n}(P)$ and the vector $\boldsymbol{\alpha} \in T_P V^{n-1}$. Rotating $\mathbf{R}^2$ about $\mathbf{n}(P)$, we obtain all normal sections of the hypersurface V^{n-1} at the point P.

For the normal section $\gamma = \mathbf{R}^2 \cap V^{n-1}$ the formula of Theorem 2 takes the form

$$k = \frac{Q(\boldsymbol{\alpha}, \boldsymbol{\alpha})}{\mathfrak{G}(\boldsymbol{\alpha}, \boldsymbol{\alpha})} = \frac{q_{ij}\alpha^i\alpha^j}{g_{ij}\alpha^i\alpha^j},$$

because $\theta = 0$. Since the curvature of a plane section (along the vector $\boldsymbol{\alpha}$) making the angle θ with the normal $\mathbf{n}(P)$ depends on θ, this function should be written as $k(\theta, \boldsymbol{\alpha})$. From the statement proved above we have $k(\theta, \boldsymbol{\alpha}) \cos\theta = k(0, \boldsymbol{\alpha})$ where $k(0, \boldsymbol{\alpha})$ is the curvature of the normal section (along $\boldsymbol{\alpha}$).

Thus, if the curvature of a normal section $k(0, \boldsymbol{\alpha})$ is known, the curvature of any plane section (along $\boldsymbol{\alpha}$) making an angle θ with $\mathbf{n}$ can be found from the relation $k(\theta, \boldsymbol{\alpha}) = \frac{1}{\cos\theta} k(0, \boldsymbol{\alpha})$.

Let us recall that in the tangent plane $T_P V^{n-1}$ there always exist principal directions $\mathbf{e}_1(P), \ldots, \mathbf{e}_{n-1}(P)$ uniquely defined if $\lambda_i \neq \lambda_j$ for $i \neq j$. Consider these "principal axes" and construct for each axis the corresponding normal plane section $\gamma_i = \mathbf{R}_i^2 \cap V^{n-1}$, where the plane $\mathbf{R}_i^2$ is spanned by $\mathbf{n}(P)$ and $\mathbf{e}_i(P)$. All the principal directions $\mathbf{e}_i(P)$ are mutually orthogonal in the Euclidean metric on $T_P V^{n-1}$. Let $k_i(P)$ be the curvature of a normal section γ_i (such a section is sometimes called a principal normal section).

Theorem 3. *The eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$ coincide in modulus with the curvatures $k_1, \ldots, k_{n-1}$ of principal normal sections.*

Proof. From Theorem 2 we have $k \cdot \cos\theta = \frac{q_{ij}\alpha^i\alpha^j}{g_{ij}\alpha^i\alpha^j}$. Since for a normal section $\theta = 0, \pi$ we have $\pm k = \frac{q_{ij}\alpha^i\alpha^j}{g_{ij}\alpha^i\alpha^j}$, where $\boldsymbol{\alpha}$ is the vector defining the normal section. Fix in $T_P V^{n-1}$ an orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$, then $g_{ij} = \delta_{ij}$, $q_{ij} = \delta_{ij}\lambda_i$, and hence

$$\pm k = \frac{\sum_{i=1}^{n-1} \lambda_i (\alpha^i)^2}{\sum_{i=1}^{n-1} (\alpha^i)^2}.$$

If $\mathbf{a} \in T_P V^{n-1}$ coincides with one of the $\mathbf{e}_i$, then $\pm k_i = \lambda_i$, which is what was required.

Let us consider in $T_P V^{n-1}$ an arbitrary vector $\boldsymbol{\alpha}$ and normal section along $\boldsymbol{\alpha}$. Let φ_i $(1 \leqslant i \leqslant n-1)$ denote the angles between $\boldsymbol{\alpha}$ and the principal directions $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ (Fig. 4.14).

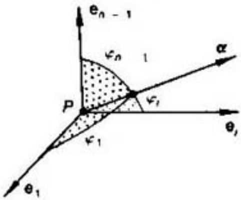

Figure 4.14

Proposition 1. *For the normal section along an arbitrary vector $\alpha \in T_PV^{n-1}$ the following relation (the Euler formula) is valid:*

$$k = k(\alpha) = \sum_{i=1}^{n-1} \lambda_i \cos^2 \varphi_i.$$

Proof. By Theorem 3 we have

$$k(\alpha) = \frac{\sum_{i=1}^{n-1} \lambda_i (\alpha^i)^2}{\sum_{i=1}^{n-1} (\alpha^i)^2} = \sum_{i=1}^{n-1} \lambda_i \left(\frac{\alpha_i}{\sqrt{\sum_{i=1}^{n-1} (\alpha^i)^2}} \right)^2 = \sum_{i=1}^{n-1} \lambda_i \cos^2 \varphi_i,$$

where

$$\cos \varphi_i = \frac{\alpha^i}{\sqrt{\sum_{i=1}^{n-1} (\alpha^i)^2}} = \frac{\alpha^i}{|\alpha|},$$

and $|\alpha|$ stands for the length of α $\left(\text{it is obvious that } \cos \varphi_i = \frac{\alpha^i}{|\alpha|}\right)$.

The Euler formula permits an analysis of the so-called "extremal" properties of the principal curvatures $\lambda_1, \ldots, \lambda_{n-1}$. Consider the curvature of the normal section $k(\alpha)$ as a function of $\alpha \in T_PV^{n-1}$. Since $k(\alpha) = (\rho, \alpha)$, where $\rho \neq 0$ is a real number, the curvature $k(\alpha)$ depends only on the direction cosines $\cos \varphi_1, \ldots, \cos \varphi_{n-1}$, where the angles $\varphi_1, \ldots, \varphi_{n-1}$ were introduced above and $\sum_{i=1}^{n-1} \cos^2 \varphi_i = 1$. Since $\{\cos \varphi_i,\ 1 \leqslant i \leqslant n-1\}$ may be considered as the coordinates of the unit vector $\alpha/|\alpha|$, by putting $x^i = \cos \varphi_i$, $1 \leqslant i \leqslant n-1$, we

may assume that $k(\alpha) = k(x^1, \ldots, x^{n-1}) = \sum_{i=1}^{n-1} \lambda_i (x^i)^2$ is a smooth function on the sphere S^{n-2} given in $T_P V^{n-1}$ by the equation $(x^1)^2 + \ldots + (x^{n-1})^2 = 1$. From the relation $k(\alpha) = k(-\alpha)$ we observe that the curvature $k(\alpha)$ is a smooth function on the projective space $\mathbf{RP}^{n-2} = S^{n-2}/\mathbf{Z}_2$ ($\mathbf{RP}^{n-2}$ is a smooth manifold). Thus, $k(\alpha)$ is a function of $(n-2)$ variables $\left(x^1, \ldots, x^{n-1}, \sum_{i=1}^{n-1} (x^i)^2 = 1\right)$. Since S^{n-2} and $\mathbf{RP}^{n-2}$ are smooth manifolds, we can introduce local coordinates $y^1, \ldots, y^{n-2}$ in a neighbourhood of each point $x \in S^{n-2}$ (for instance, if $x^{n-1} \geqslant 0$, we may assume $x^1, \ldots, x^{n-2}$ to be coordinates on S^{n-2}). Let us call $x_0 \in S^{n-2}$ a *critical point* for the function $f(x)$ given on S^{n-2}, if $\left.\frac{\partial f}{\partial y^i}\right|_{x_0} = 0$, $1 \leqslant i \leqslant n-2$. The equality $\frac{\partial f}{\partial y^{i'}} = \frac{\partial f}{\partial y^i} \frac{\partial y^i}{\partial y^{i'}}$ implies that if a point x_0 is critical in one coordinate system, say $y^1, \ldots, y^{n-2}$, it will also be critical in any other coordinate system. Thus, the definition of the critical point is invariant.

Question: What are critical points of the smooth function $k(\alpha) = \sum_{i=1}^{n-1} \lambda_i (x^i)^2$ on the sphere $S^{n-2} \subset T_P V^{n-1}$? Furthermore, what values does the curvature $k(\alpha)$ take at these critical points (i.e. along the directions of α of corresponding to the critical points on S^{n-2})?

Theorem 4. *The critical points of the curvature function $k(\alpha)$ on the sphere $S^{n-2} \subset T_P V^{n-1}$, are exactly the points $\pm e_i$, $1 \leqslant i \leqslant n-1$ (i.e. the ends of the vectors $\pm \mathbf{e}_i$) when $\lambda_i \neq \lambda_j$ for $i \neq j$. At these points $k(\alpha)$ has the values λ_i, $1 \leqslant i \leqslant n-1$. In this sense, the principal curvatures are the extremal values of the curvature function $k(\alpha)$. If some of the eigenvalues $\{\lambda_i\}$ coincide, the theorem will read: the critical points of the function $k(\alpha)$ are exactly the ends of all eigenvectors of the form Q.*

Proof. Let us consider an arbitrary bilinear symmetric form $B(\mathbf{x}, \mathbf{y})$ defined on the vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^{n-1}$ (this can be, for example, the second fundamental form $Q(\mathbf{x}, \mathbf{y}) = q_{ij} x^i y^j$). Construct the function $f(x) = \frac{B(\mathbf{x}, \mathbf{x})}{|\mathbf{x}|^2}$, $x \in S^{n-2}$, and find all its critical points. Obviously, $f(x) = B(\mathbf{x}, \mathbf{x})$ if $|\mathbf{x}| = 1$. That $x_0 \in S^{n-2}$ is a critical point for $f(x)$ means that $\left.\frac{df}{d\mathbf{b}}\right|_{x_0} = 0$ for any vector $\mathbf{b}$ tangent to S^{n-2} at x_0, where $\left.\frac{df}{d\mathbf{b}}\right|_{x_0}$ denotes the derivative of $f(x)$ with respect to $\mathbf{b}$, i.e. $\left.\frac{df}{d\mathbf{b}}\right|_{x_0} = \left.\frac{\partial f(y^1, \ldots, y^{n-2})}{\partial y^i}\right|_{x_0} \cdot b^i$, where y^i are

local coordinates in some neighbourhood of x_0. Using an equivalent definition of this derivative, we have $\frac{df}{d\mathbf{b}}\Big|_{x_0} = \frac{df(\gamma(t))}{dt}\Big|_{t=0}$, where $\gamma(t)$ is an arbitrary curve entirely lying on the sphere S^{n-2}, and $\gamma(0) = x_0$, $\frac{d\gamma(t)}{dt}\Big|_{t=0} = \mathbf{b}$ (see Fig. 4.15). It follows that $\frac{df(\gamma(t))}{dt}\Big|_{t=0} = \frac{d}{dt} B(\gamma(t), \gamma(t))\Big|_{t=0}$. Since the scalar product $\langle\, ,\rangle$

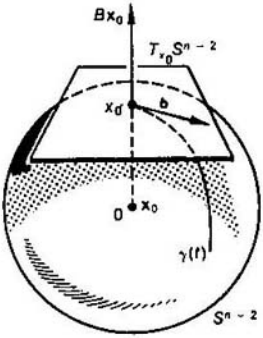

Figure 4.15

is fixed in $\mathbf{R}^{n-1} \supset S^{n-2}$, we can associate bijectively with the form $B(\mathbf{x}, \mathbf{y})$ a symmetric operator B: $\mathbf{R}^{n-1} \to \mathbf{R}^{n-1}$ such that $B(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, B\mathbf{y}\rangle$, i.e. $f(x) = B(\mathbf{x}, \mathbf{x}) = \langle \mathbf{x}, B\mathbf{x}\rangle$, where $x \in S^{n-1}$ is the end of the vector $\mathbf{x}$ pointing from 0 to $x \in S^{n-2}$. Hence,

$$
\begin{aligned}
\frac{df(x)}{d\mathbf{b}}\Big|_{x_0} &= \frac{d}{dt} B(\gamma(t), \gamma(t))\Big|_{t=0} = \frac{d}{dt}\langle \gamma(t), B\gamma(t)\rangle \\
&= \left\langle \frac{d}{dt}\gamma(t), B\gamma(t)\right\rangle + \left\langle \gamma(t), B\frac{d}{dt}\gamma(t)\right\rangle \\
&= \left\langle \frac{d}{dt}\gamma(t), B\gamma(t)\right\rangle + \left\langle B\gamma(t), \frac{d}{dt}\gamma(t)\right\rangle \\
&= 2\left\langle \frac{d}{dt}(\gamma(t), B\gamma(t)\right\rangle.
\end{aligned}
$$

Since $\frac{d}{dt}\gamma(t)\Big|_{t=0} = \mathbf{b}$, we have $\frac{d}{d\mathbf{b}} f(x)\Big|_{\mathbf{x}_0} = 2\langle \mathbf{b}, B\mathbf{x}_0\rangle = 0$, i.e. $\langle \mathbf{b}, B\mathbf{x}_0\rangle = 0$, and hence for any $\mathbf{b} \in T_{x_0}S^{n-2}$ the vector $B\mathbf{x}_0$ is orthogonal to $T_{x_0}S^{n-2}$ (Fig. 4.15). Thus, $B\mathbf{x}_0$ is collinear to $\mathbf{x}_0$, i.e. $\lambda(x_0)\cdot\mathbf{x}_0 = Bx_0$, where $\lambda(x_0) \neq 0$ is a (real) eigenvalue. Hence, the critical points of $f(x) = B(\mathbf{x}, \mathbf{x})$ on S^{n-2} are only those points

which are the ends of the unit eigenvectors of the form B. Since an orthogonal basis can be chosen from the eigenvectors of B, $\lambda(x_0)$ coincides with the eigenvalue for $\mathbf{x}_0$. The theorem is proved.

Let $n = 3$ and V^2 be embedded in $\mathbf{R}^3$. Then the Euler formula becomes $k(\alpha) = \lambda_1 \cos^2 \varphi_1 + \lambda_2 \cos^2 \varphi_2$, where $\cos^2 \varphi_1 + \cos^2 \varphi_2 = 1$; suppose $\lambda_1 \geqslant \lambda_2$, then $k(\alpha) = (\lambda_1 - \lambda_2) \cos^2 \varphi_1 + \lambda_2$. Apparently, $k(\alpha)$ has a minimum for $\cos^2 \varphi_1 = 0$: $k(\alpha) = \lambda_2$, and a maximum for $\cos^2 \varphi_1 = 1$: $k(\alpha) = \lambda_1$. If $\lambda_1 = \lambda_2$, then $k(\alpha) = \lambda\ (= \lambda_1 = \lambda_2)$ (see Fig. 4.16).

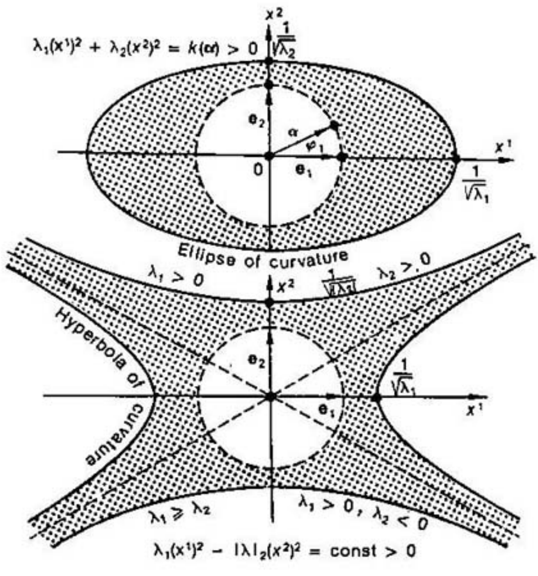

Figure 4.16

Geometric interpretation of the principal curvatures leads to an elegant geometric pattern, the so-called quadric of normal curvatures: $\lambda_1 \cos^2 \varphi_1 + \ldots + \lambda_{n-1} \cos^2 \varphi_{n-1} = \text{const}$. The directions corresponding to the extremal (critical) values of the curvature coincide with the principal axes of this surface. If all principal curvatures $\lambda_1, \ldots, \lambda_{n-1}$ are non-negative, the quadric is an ellipsoid with the principal semiaxes $1/\sqrt{\lambda_i}$, $1 \leqslant i \leqslant n-1$, i.e. the ith principal

semiaxis is equal to $\sqrt{r_i}$, where r_i is the corresponding curvature radius of the ith normal section $\gamma_i = \mathbf{R}^2 \cap V^{n-1}$.

Let us analyse the case of a two-dimensional surface $V^2 \subset \mathbf{R}^3$. It appears that the Gaussian curvature $K(P)$ determines some important local characteristics of a two-dimensional surface. We recall that $K(P) = \lambda_1 \cdot \lambda_2$, and three cases are thus possible:

(a) $K > 0$, (b) $K < 0$, (c) $K = 0$, whence

$$\text{(a)} \begin{cases} \lambda_1 > 0 \\ \lambda_2 > 0 \end{cases} \text{or} \quad \begin{cases} \lambda_1 < 0 \\ \lambda_2 < 0 \end{cases}; \quad \text{(b)} \begin{cases} \lambda_1 > 0 \\ \lambda_2 < 0 \end{cases} \text{or} \quad \begin{cases} \lambda_1 < 0 \\ \lambda_2 > 0 \end{cases};$$

$$\text{(c)} \begin{cases} \lambda_1 = 0 \\ \lambda_2 \neq 0 \end{cases} \text{or} \quad \begin{cases} \lambda_1 \neq 0 \\ \lambda_2 = 0 \end{cases} \text{or} \quad \lambda_1 = \lambda_2 = 0.$$

It is sufficient to consider only cases: (a) $\lambda_1 > 0, \lambda_2 > 0$, (b) $\lambda_1 > 0$, $\lambda_2 < 0$, (c) $\lambda_1 = 0$, $\lambda_2 \neq 0$, $\lambda_1 = \lambda_2 = 0$.

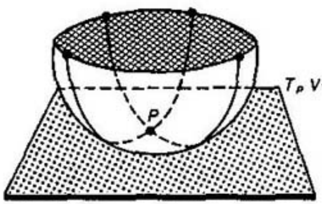

Figure 4.17

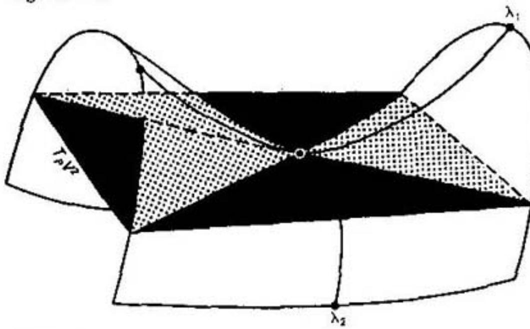

Figure 4.18

In case (a) the shape of the surface V^2 near the point P is (locally) as that in Fig. 4.17. Here V^2 lies on the side of the tangent plane at

P. In case (b) the shape of V^2 near P (locally) is shown in Fig. 4.18. Here V^2 lies on both sides of $T_P V^2$. In this case the point P is called a *saddle point* (or simply a saddle), sometimes it is called a *hyperbol-*

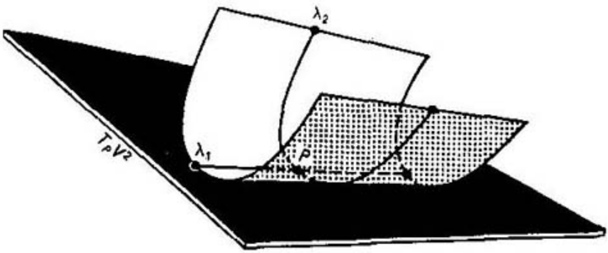

Figure 4.19

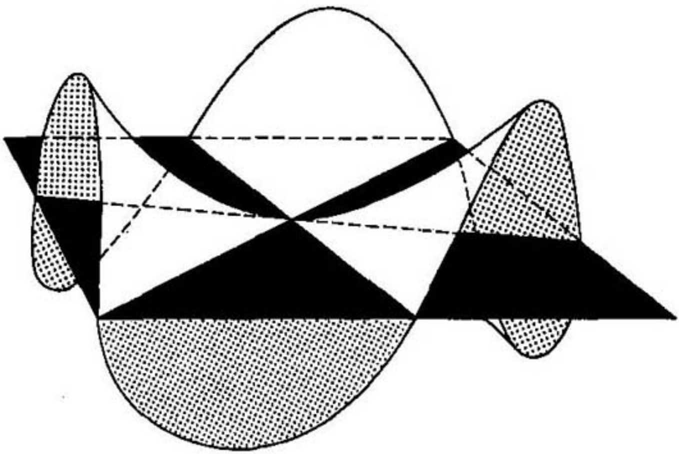

Figure 4.20

ic point. In the third case (c), $\lambda_2 \neq 0$, $\lambda_1 = 0$, the surface V^2 near P is (locally) as in Fig. 4.19. Since $\lambda_1 \neq \lambda_2$, V^2 has at this point two orthogonal principal directions, $\mathbf{e}_1$ and $\mathbf{e}_2$. One should not believe that near the point P, at which $\lambda_1 = \lambda_2 = 0$, the surface V^2 has the

local structure of a plane (from the metric point of view). The fact is that if the second fundamental form Q vanishes at P (i.e. when $\lambda_1 = \lambda_2 = 0$) the local structure of V^2 near P is not quadratic. As an example of V^2, we consider the graph $z = f(x, y) = \operatorname{Re}(x + iy)^3 = x^3 - 3xy^3$ (see Fig. 4.20, the so-called monkey saddle). The first fundamental form $\mathfrak{G}$ at the point $(x = 0, y = 0)$ is $g_{ij} = \delta_{ij}$.

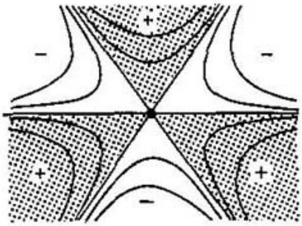

Figure 4.21

The second fundamental form Q at the point $(0, 0)$ is degenerate, since $\lambda_1 = \lambda_2 = 0$. Here $Q = (q_{ij})$, where $q_{ij} = f_{x^i x^j}$, i.e. $Q = 6\begin{pmatrix} x & -y \\ -y & -x \end{pmatrix}$; hence $Q = 0$ at the point $(0, 0)$. The family of level lines of the function $f(x, y)$ at $(0, 0)$ is shown in Fig. 4.21.

4.2.4. THE GAUSSIAN AND MEAN CURVATURES OF A TWO-DIMENSIONAL SURFACE

Now we shall find an explicit form of the Gaussian and mean curvatures of a two-dimensional surface defined explicitly. Let $V^2 \subset \mathbb{R}^3$ be given as the graph $z = f(x, y)$ where (x, y, z) are Cartesian coordinates in $\mathbb{R}^3$. Let $f_x = f_y = f(0, 0) = 0$, then the coordinate plane (x, y) is tangent to V^2 at the point $(0, 0)$. Since

$$\mathfrak{G}\,|_{(0,0)} = (\delta_{ij}) = E, \quad Q\,|_{(0,0)} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

we have $K = \det Q = f_{xx} f_{yy} - f_{xy}^2$, $H = \operatorname{spur} Q = f_{xx} + f_{yy} = \lambda_1 + \lambda_2$. The curvatures H and K at points distinct from $(0, 0)$ can be calculated as follows. Let $\mathfrak{G} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$ and $Q = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$, then

$\mathfrak{G}^{-1} = \frac{1}{g}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}$, where $g = \det \mathfrak{G}$. Hence,

$$\mathfrak{G}^{-1}Q = \frac{1}{g}\begin{pmatrix} GL-FM & GM-FN \\ -FL+EM & -FM+EN \end{pmatrix},$$

$$H = \operatorname{spur}(\mathfrak{G}^{-1}Q) = \frac{1}{g}(GL-2FM+EN),$$

$$K = \det(\mathfrak{G}^{-1}Q) = \frac{LN-M^2}{EG-F^2}.$$

If $z = f(x, y)$, then $ds^2 = (1 + f_x^2)\,dx^2 + 2f_xf_y\,dx\,dy + (1 + f_y^2)\,dy^2$. Since the radius vector of V^2 given by the graph $z = f(x, y)$ is of the form $\mathbf{r}(x, y) = (x, y, f(x, y))$, we have

$$\mathbf{r}_{xx} = (0, 0, f_{xx}), \quad \mathbf{r}_{xy} = (0, 0, f_{xy}), \ \mathbf{r}_{yy} = (0, 0, f_{yy}),$$

$$\mathbf{n} = \frac{\operatorname{grad}(z-f)}{|\operatorname{grad}(z-f)|} = \frac{(-f_x, -f_y, 1)}{\sqrt{1+f_x^2+f_y^2}},$$

$$L = \frac{f_{xx}}{\sqrt{1+f_x^2+f_y^2}}, \quad M = \frac{f_{xy}}{\sqrt{1+f_x^2+f_y^2}}; \quad N = \frac{f_{yy}}{\sqrt{1+f_x^2+f_y^2}},$$

whence

$$H = \frac{GL-2MF+EN}{1+f_x^2+f_y^2} = \frac{(1+f_y^2)f_{xx}-2f_xf_yf_{xy}+(1+f_x^2)f_{yy}}{(1+f_x^2+f_y^2)^{3/2}},$$

$$K = \frac{f_{xx}f_{yy}-f_{xy}^2}{(1+f_x^2+f_y^2)^2}.$$

The Gaussian and mean curvatures are scalar functions defined at each point of a surface, and they are also invariants of the surface, in particular, these curvatures do not depend on the choice of local coordinates.

Let us calculate the Gaussian and mean curvatures of a standardly embedded sphere $S^2 \subset \mathbf{R}^3$. Since any normal section of S^2 at an arbitrary point P is an equator, $\lambda_1 = \lambda_2 = 1/R$, where R is the sphere radius. Hence, the curvature of any normal section is $\lambda = 1/R$ and therefore $K = 1/R^2$, $H = 2/R$; in particular, the Gaussian and mean curvatures are constant.

The Gaussian and mean curvatures of a two-dimensional plane are equal to zero.

Definition. A two-dimensional surface $V^2 \subset \mathbf{R}^3$ is called the *surface of constant curvature* if its Gaussian curvature is constant.

For example, a standard sphere S^2 and a Euclidean plane are manifolds of constant curvature.

Definition. A two-dimensional surface $V^2 \subset \mathbf{R}^3$ is called the *surface of positive, zero or negative curvature* if the Gaussian curvature at all points of this surface is positive, zero or negative, respectively.

A standard two-dimensional sphere is a manifold of positive (constant) curvature.

Problem. Prove that the surface $V^2 \subset \mathbf{R}^3$ given by the equation $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$ (ellipsoid) is the surface of strictly positive curvature, provided the semiaxes a, b, and c are not equal to 0 and ∞. *Hint*: the ellipsoid can be written parametrically as

$$x=a\cos\theta\cos\varphi,\quad y=b\cos\theta\sin\varphi,\quad z=c\sin\theta.$$

A Euclidean two-dimensional plane is a manifold of zero constant curvature. The graph $z=x^2-y^2$ is an example of a manifold of negative curvature; obviously, $K=\frac{f_{xx}f_{yy}-f_{xy}^2}{(1+f_x^2+f_y^2)^2}=\frac{-4}{(1+4(x^2+y^2))^2}<0$. In this example the Gaussian curvature is a variable function. It would be desirable, by analogy with a surface of positive or zero curvature, to construct a manifold of constant negative curvature, $V^2\subset\mathbf{R}^3$. Now we shall give an example of such a surface, thereby proving the following **statement**: *in a three-dimensional Euclidean space there exist (locally) surfaces of constant positive, zero or negative Gaussian curvature.*

Let us consider a smooth curve γ located in the first quadrant of the plane (x, y) and having the following property: the length of the tangent from the point where it touches the curve to the point where it intersects the x-axis is constant and equal to a (Fig. 4.22).

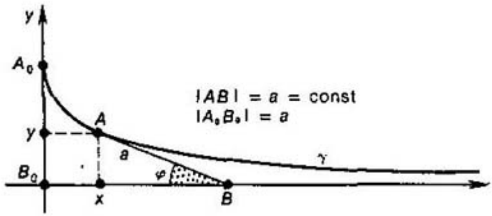

Figure 4.22

As the point A is moving along the curve γ, the point B slides along the x-axis and the segment AB has the constant length, a. Such a curve can be obtained if the points A and B are connected by an inextensible string of length a (the initial positions of the points are A_0 and B_0, see Fig. 4.22), and B is made to move along the x-axis. Then A will describe the curve γ which touches the y-axis at A_0 and has the x-axis as an asymptote. The points A and B are assumed to move along the plane without friction. In this case the velocity vector

is always directed along the string connecting A and B, i.e. the motion of B is uniquely defined. Let us derive a differential equation of the curve γ. From the triangle ABx we have (see Fig. 4.22): $\tan\varphi = -y'_x$, where $y = y(x)$ is the graph of γ, and $a \cdot \sin\varphi = y$, whence

$$\sin\varphi = \frac{-y'_x}{\sqrt{1+(y'_x)^2}}, \text{ i.e. } \frac{ay'_x}{\sqrt{1+(y'_x)^2}} = -y; \quad x'_y = \frac{\sqrt{a^2-y^2}}{y},$$

where $x = x(y)$ is the graph of γ. Hence,

$$x(y) = -\int_a^y \frac{1}{y}\sqrt{a^2-y^2}\,dy = -a^2\int_a^y \frac{dy}{y\sqrt{a^2-y^2}} + \int_a^y \frac{y\,dy}{\sqrt{a^2-y^2}}$$

$$= -a^2\int_a^y \frac{dy}{y\sqrt{a^2-y^2}} - \sqrt{a^2-y^2}$$

$$= -\sqrt{a^2-y^2} + \frac{a}{2}\ln\left(\frac{a+\sqrt{a^2-y^2}}{a-\sqrt{a^2-y^2}}\right).$$

Thus, we have derived the explicit expression for $x = x(y)$. Let us consider the surface which is obtained if the curve γ is rotated about the x-axis (see Fig. 4.23). Such a surface V^2 is called the

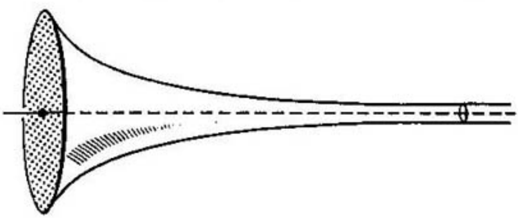

Figure 4.23. Beltrami surface

Beltrami surface or *pseudosphere*; the latter term will be explained below. To find the Gaussian curvature of the Beltrami surface, we have to calculate the Gaussian curvature of a surface of rotation. Let us solve this problem in the general form.

Let us consider in $\mathbf{R}^3(x, y, z)$ the surface V^2 obtained by the rotation about the x-axis of a smooth curve $x = x(y)$ (the generator) lying in the plane Oxy. A coordinate network, parallels and meridians, is obtained on V^2, and this network is orthogonal because at

each point the coordinate lines intersect at right angles (see Fig. 4.24).

Lemma 4. *At each point of a surface of rotation the principal directions, i.e. the directions corresponding to the principal curvatures λ_1 and λ_2, may always be considered as coinciding with the directions of the meridian and parallel through this point.*

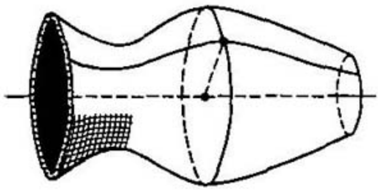

Figure 4.24

Remark. The words "may be considered" have the following meaning: if $\lambda_1 \neq \lambda_2$, the principal directions are defined uniquely and coincide with the directions of the meridian and parallel; if $\lambda_1 = \lambda_2$, any direction is principal, including, as a particular case, the orthogonal directions of the meridian and parallel.

Proof. As we know, the principal directions in $T_P V^2$ are only those orthogonal bases $\mathbf{e}_1$, $\mathbf{e}_2$ in which the forms $\mathfrak{G}$ and Q are both diagonal. Apparently, the first fundamental form $\mathfrak{G}$ is diagonal in the coordinate system generated by the meridians and parallels (see, for instance, the calculation of $\mathfrak{G}$ for a surface of rotation). It remains to prove therefore that the second fundamental form Q is also diagonal in this coordinate system. Let (u, v) be the coordinates on V^2 which generate a coordinate network, meridians and parallels. We should prove that

$$M = \langle \mathbf{r}_{uv}, \mathbf{n} \rangle \equiv 0,$$

where $\mathbf{r}(u, v)$ is the radius vector of V^2, $\mathbf{n}$ is the normal to V^2, and $Q = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$ is the second fundamental form. Consider cylindrical coordinates (r, φ, x) in $\mathbf{R}^3$ and let V^2 be given by the generator $r = r(x)$. Then the radius vector $\mathbf{r}(x, \varphi)$ of V^2 is of the form $\mathbf{r}(x, \varphi) = (x, r(x) \cos \varphi, r(x) \sin \varphi)$ (see Fig. 4.25), whence $\mathbf{r}_{x\varphi} = (0, -r' \sin \varphi, r' \cos \varphi)$. The normal to V^2 is given by $\mathbf{n} = (r', -\cos \varphi, -\sin \varphi)/\sqrt{1 + (r')^2}$. Obviously, $\langle \mathbf{r}_{x\varphi}, \mathbf{n} \rangle = 0$, so that $Q = \begin{pmatrix} L & 0 \\ 0 & N \end{pmatrix}$, which is what was to be proved.

Lemma. 5. *The Gaussian curvature $K(P)$ of a surface of rotation V^2 at a point $P \in V^2$ is of the form* $|K| = \frac{|r''|}{r(1+(r')^2)^2}$, *where $r = r(x)$ is the equation of the generator of V^2 in cylindrical coordinates.*

Remark. In other words, $K = \lambda_1\lambda_2$, where $\frac{1}{r\sqrt{1+(r')^2}} = \lambda_1$ is the curvature of the normal section along the parallel at the point P, and $\lambda_2 = k(x)$ is the curvature of the plane curve $y = y(x)$ at

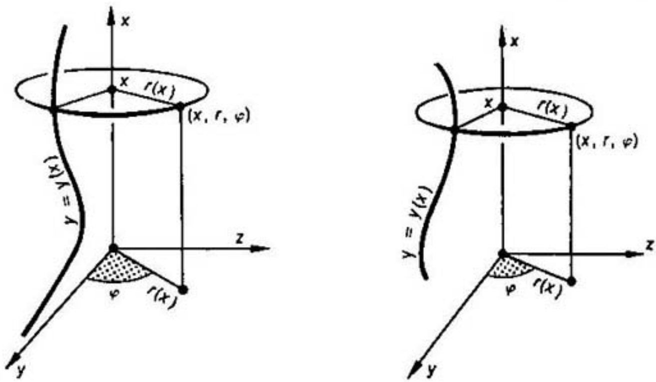

Figure 4.25

a point x, i.e. the curvature of the meridian. The normal section along a parallel does not generally coincide with the parallel itself (the formula for λ_1 will be proved below).

Proof. According to Lemma 4, the principal directions coincide with the directions of the meridian and parallel through the point P and therefore $K(P) = \lambda_1\lambda_2$, where λ_2 and λ_1 are the curvatures of plane curves—the meridian and the normal section along the parallel, respectively. Since the parallel is a circle, we shall use the corresponding formula to find the curvature of the normal section. Since the meridian coincides with the generator $r = r(x)$, it follows from the Frenet formulas for plane curves that $|\lambda_2(x)| = \frac{|r''|}{(1+(r')^2)^{3/2}}$. We now find $\lambda_1(x)$ for the normal section along the parallel at the point $P = (x, r(x))$. Consider the parallel as a plane section of V^2; then the curvature $k(\theta, \alpha)$ of this section (along α defined by the parallel) is equal to $1/r(x)$, since the radius of the circle (parallel)

is $r(x)$. Here θ is the angle between the normal to V^2 and the vector $\mathbf{m}$ lying in the plane of the parallel (Fig. 4.26). Recall that

$$k(\theta, \alpha) = \frac{1}{\cos\theta} \cdot k(0, \alpha)$$

(see above). Here $k(0, \alpha) = \frac{\cos\theta}{2} = \cos\theta \cdot k(\theta, \alpha)$ is the curvature in question, λ_1. Thus, it remains to find $\cos\theta$, where θ is the angle

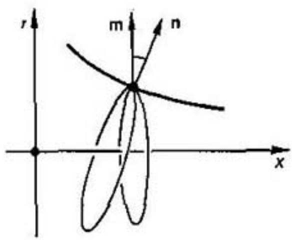

Figure 4.26

between the unit vectors $\mathbf{m}$ and $\mathbf{n}$. Since in the plane xOr the vector $\mathbf{m}$ has the coordinates $(0, 1)$ and the vector $\mathbf{n}$ the coordinates $\frac{1}{\sqrt{1+(r')^2}}(-r', 1)$, we have $\cos\theta = \frac{1}{\sqrt{1+(r')^2}}$ and

$\lambda_1(x) = \frac{1}{r\sqrt{1+(r')^2}}$, whence

$$|K| = \frac{1}{r\sqrt{1+(r')^2}} \cdot \frac{|r''|}{(1+(r')^2)^{3/2}} = \frac{|r''|}{r(1+(r')^2)^2}.$$

The lemma is proved.

Lemma 6. *The Beltrami surface is a manifold of constant (strictly) negative curvature.*

Proof. Since the Beltrami surface is a surface of rotation, we can use the formula of Lemma 5 to calculate the Gaussian curvature. The function $y = y(x)$ is the inverse of the function

$$x = x(y) = -\sqrt{a^2-y^2} + \frac{a}{2}\ln\left(\frac{a+\sqrt{a^2-y^2}}{a-\sqrt{a^2-y^2}}\right)$$

obtained above. As was already calculated,

$$x'_y = \frac{-\sqrt{a^2-y^2}}{y}, \quad \text{whence } x'' = \frac{a^2}{r^2\sqrt{a^2-r^2}},$$

and consequently

$$K = \frac{r''}{r(1+(r')^2)^2} = \frac{-x''}{(x')^3 r\left(1+\frac{1}{(x')^2}\right)^2} = \frac{-x''x'}{r(1+(x')^2)^2} = -\frac{1}{a^2}.$$

The minus sign appears because the curve $y = y(x)$ is convex downward and therefore the eigenvalues λ_1, λ_2 have opposite signs relative to any normal $\mathbf{n}(P)$. Thus, $K = -1/a^2$, and this completes the proof of the lemma.

We have constructed in $\mathbf{R}^3$, at least locally, the surfaces of constant positive, zero, and negative Gaussian curvature. The manifold of constant positive curvature (a sphere) is compact and closed (without boundary); the manifold of zero curvature (a plane or a cone formed

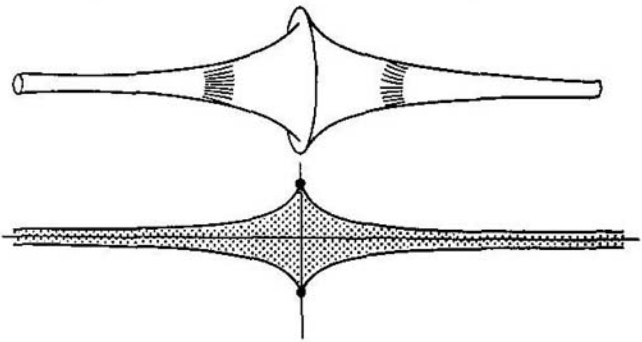

Figure 4.27. Section of the Beltrami surface

by straight lines emerging from a single point, finite or infinite, and moving along an arbitrary smooth plane curve γ in $\mathbf{R}^3$) is non-compact and open (without boundary). The manifold of constant negative curvature just considered differs from these two examples in that the surface of negative curvature is not a closed manifold and cannot be extended to infinity. This surface (see Fig. 4.23) has a boundary, the circle of radius a with centre at the origin O, and it can be shown (we shall not concentrate on this point) that the surface cannot be extended beyond this circle in such a manner that the condition $K(P) = -\frac{1}{a^2} < 0$ is satisfied.

This surface is usually "completed" by adjoining a surface symmetric to the initial one about the plane yOz (Fig. 4.27). The resultant surface has a "cusp circle" where the surface is not a smooth submanifold in $\mathbf{R}^3$. This circle formed by singular points is not acciden-

tal. It seems at first sight that these singularities could have been avoided if the "funnel" constructed above had been extended beyond the circle of radius a, as in Fig. 4.28. (We recall that at the point A_0 on the y-axis the graph of the generator touches this axis.) Once such an extension has been carried out, we find that the surface of rotation thus obtained is not a surface of negative curvature: the part swept out by the arc A_0C (see Fig. 4.28) has positive curvature, since the arc A_0C is convex upward (unlike the arc A_0A).

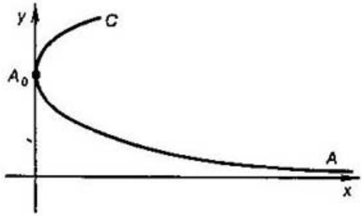

Figure 4.28

Hence, it is doubtful whether the Beltrami surface can be smoothly extended beyond the circle described by the point A_0 around the x-axis such that the surface curvature remains negative. Thus, it is rather difficult to construct in $\mathbf{R}^3$ a closed, compact or non-compact, but infinite in all directions manifold of constant negative curvature. Here we can easily see the difference between a surface of constant positive curvature and a surface of constant negative curvature (surfaces of variable negative curvature extending in $\mathbf{R}^3$ infinitely in all directions do exist: for example, the surface of hyperbolic paraboloid $z = x^2 - y^2$ considered above; here the Gaussian curvature is negative and vanishes at infinity if, for example, $x^2 + y^2 \to \infty$). Indeed, a standard two-dimensional sphere has a constant positive curvature in the metric induced by this embedding.

To get a deeper insight in the situation under discussion, we shall find the induced Riemannian metric on the Beltrami surface embedded in $\mathbf{R}^3$. Introduce in $\mathbf{R}^3$ cylindrical coordinates (x, r, φ) where $x = x$, $y = r \cos \varphi$, $z = r \sin \varphi$ (i.e. the x-axis is the rotation axis). Then the metric induced on the surface of rotation with the generator $x = x(r)$ is of the form

$$ds^2 = (dx(r))^2 + dr^2 + r^2 d\varphi^2 = (1 + (x_r')^2) dr^2 + r^2 d\varphi^2.$$

In our example $x_r' = -\frac{\sqrt{a^2 - r^2}}{r}$ (see above), i.e. $ds^2 = \frac{a^2 dr^2}{r^2} + r^2 d\varphi^2$.

Statement 1. *The Riemannian metric induced on the Beltrami surface by the ambient Euclidean metric is a Lobachevskian metric.*

Proof. By virtue of the transformation $u=\varphi$, $v=1/r$ we have $ds^2=\frac{v^2}{v^4}\,dv^2+\frac{du^2}{v^2}=\frac{du^2+dv^2}{v^2}$, which, apparently, proves the statement. Thus, the Beltrami surface is isometric (locally) to a Loba-

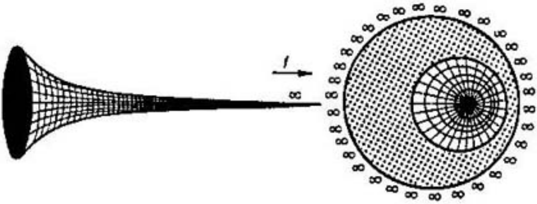

Figure 4.29

chevskian plane, i.e. we have constructed an isometric embedding of a domain of a Lobachevskian plane in a three-dimensional Euclidean space. A question arises: which part of the Lobachevskian plane admits isometric embedding in $\mathbf{R}^3$ (as the Beltrami surface)? Note first of all that the entire Beltrami surface is not isometric to any piece of the Lobachevskian plane. Indeed, the Beltrami surface is

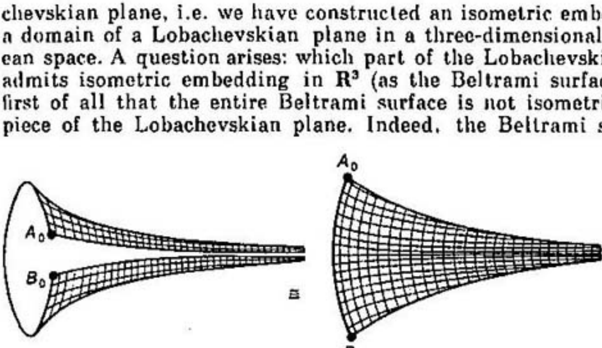

Figure 4.30

homeomorphic to a disk with a punctured point; hence, if this ring were homeomorphic to a domain in the Lobachevskian plane, an infinitely distant point of the Beltrami surface would be mapped into a finite point on the Lobachevskian plane (see Fig. 4.29). This contradicts the fact that an infinitely distant point of the Beltrami surface is infinitely removed from the "funnel" neck.

The real situation is, however, that the Beltrami surface cut along any of its generators is mapped isometrically onto a domain of the Lobachevskian plane (see Fig. 4.30). The corresponding domain

(in the Poincaré model) is shown in Fig. 4.31 as (∞, A_0, B_0); it lies between two parallel "straight" lines (in the Lobachevskian sense) emerging from the point ∞ on the absolute and the arc A_0B_0 which

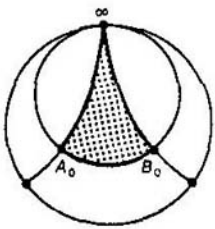

Figure 4.31

is a part, of length 2π, of the circle (in the Euclidean sense) tangent to the point ∞ on the absolute.

Thus, the domain (∞, A_0, B_0) is an infinite strip lying between two parallel straight lines and bounded from one end by the arc A_0B_0.

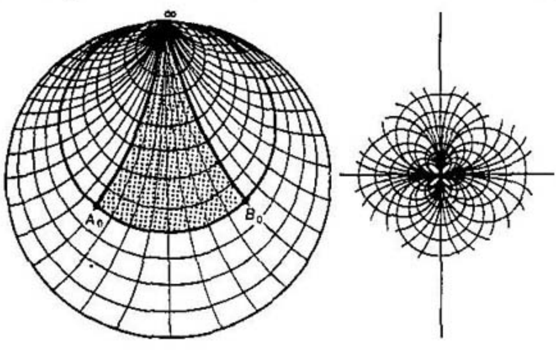

Figure 4.32

Let us consider, in the Poincaré model, two families of coordinate lines that form an orthogonal network (both in the Euclidean sense and in the sense of the Lobachevskian metric, since the two metrics differ only by a conformal factor). One family is the set of straight lines (in the Lobachevskian sense) emerging from the point ∞ on the absolute, i.e. the family of circular arcs (in the Euclidean sense) meeting the absolute at right angles. The other family represents

the set of circles tangent to the absolute at the point ∞ (see Fig. 4.32). This family of trajectories coincides with the set of level lines for the real and imaginary parts of the complex-analytic function $f(z) = 1/z$, i.e. $\operatorname{Re} f(z) = \frac{1}{r} \cos \varphi$, $\operatorname{Im} f(z) = \frac{-1}{r} \sin \varphi$, where $z = re^{i\varphi}$. The equations of these lines are of the form

$$\frac{1}{r} \cos \varphi = \text{const}_1, \quad \frac{1}{r} \sin \varphi = \text{const}_2.$$

This coordinate network is orthogonal. Whereas one of the families consists of straight lines in the Lobachevskian geometry, the trajectories of the second family are not such straight lines. These trajectories are uniquely characterized by the property that all the "perpendiculars" drawn from points of a trajectory are parallel to one another and intersect at the point ∞ on the absolute (recall that the absolute does not belong to the Lobachevskian plane) (see Fig. 4.32). It is a simple matter to prove (verify!) that any two trajectories of the second family are congruent, that is, they can be transformed into each other by an isometry of the Lobachevskian plane (i.e. by a diffeomorphism preserving the Riemannian metric). Moreover, this isometry can be chosen in the form of the homographic function $w = \frac{az+b}{cz+d}$, where a, b, c, and d are appropriate complex numbers. But we shall not need this fact below.

Let us consider an arbitrary trajectory of the second family (a circle tangent to the absolute at the point ∞) and a pair of points on it, A_0 and B_0, spaced at a distance of 2π (we assume that $a = 1$ and the radius of the Euclidean Poincaré model is also equal to unity). Then the strip lying between two perpendiculars (A_0, ∞) and (B_0, ∞) drawn at A_0 and B_0 is isometric to the Beltrami surface cut along a meridian, and the orthogonal network of meridians and parallels on the Beltrami surface is transformed into an orthogonal network of trajectories of the first and second families on the Poincaré model in the strip (∞, A_0, B_0) (see Fig. 4.33). On a Lobachevskian plane (as well as on a Euclidean plane) there always exists reflection about an arbitrary straight line; in particular, reflecting the strip (∞, A_0, B_0) about the straight line (∞, A_0), we obtain a new strip isometric to the initial one (∞, A_0, B_0) and realized as a cut of the Beltrami surface in $\mathbf{R}^3$. Reflecting then this new strip (∞, A_1, A_0) about the straight line (∞, A_1), we obtain the strip (∞, A_2, A_1) with the same properties, etc. Reflection about a straight line on a Lobachevskian plane is an isometry, so that a trajectory of the second family through the points (A_0, B_0) is transformed into itself because any isometry preserving the point ∞ transforms trajectories of the second family into trajectories of the same family. All segments (A_k, A_{k-1}), $1 \leqslant k < \infty$, have the same length, 2π. Exactly

the same procedure leads to the strips (∞, B_k, B_{k-1}), $1 \leqslant k < \infty$, having analogous properties. Thus, the entire disk D^2 bounded by a trajectory of the second family (a circle) is decomposed into infi-

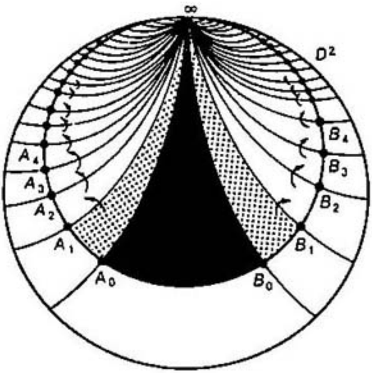

Figure 4.33

nitely many strips converging at the point ∞ on the absolute. We can now construct the mapping of the entire disk D^2 onto a Beltrami funnel (without a cut) in $\mathbf{R}^3$ such that each strip of the type (∞, A_k, A_{k-1}), (∞, A_0, B_0), and (∞, B_k, B_{k-1}), $1 \leqslant k < \infty$, is mapped

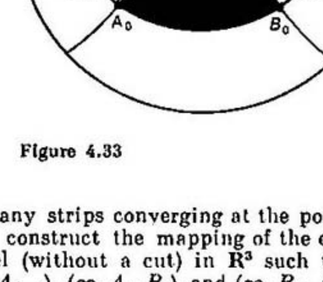

Figure 4.34

isometrically onto the Beltrami surface, the disk D^2 being wound infinitely many times onto the Beltrami surface (see Fig. 4.34).

Question: Can the entire Lobachevskian plane, and not only the strip (∞, A_0, B_0) (or any other strip isometric to the former), be realized in $\mathbf{R}^3$ as a two-dimensional surface of constant negative cur-

vature? As was shown by Hilbert, the answer is no. The following question also arises: Can we construct in $\mathbf{R}^3$ a complete two-dimensional surface of negative (variable, in general) curvature from zero by a negative number? It turns out that such a surface does not exist either. The proof of this fact is not trivial and belongs to N. V. Efimov.

This is one of the fundamental distinctions between the metrics of positive and negative curvature. Note that the induced metric on the Beltrami surface coincides with the Lobachevskian metric. In this case, the Lobachevskian metric can be defined "abstractly", irrespective of the embedding in a Euclidean space, and because of the above coincidence this metric has a Gaussian curvature. Although the initial definition of the Gaussian curvature relied upon the embedding $V^{n-1} \to \mathbf{R}^n$ and, hence, the Gaussian curvature depended on the first and second fundamental forms, it appears now that this curvature seems to be independent of the second form, i.e. it is only a function of the Riemannian metric $\mathfrak{G}$; in particular, the curvature K remains unchanged under an isometry of V^{n-1} in $\mathbf{R}^n$. This assertion, which permits the Gaussian curvature to be expressed only in terms of the Riemannian metric, will be proved below.

Let us now consider the mean curvature of a surface $V^2 \subset \mathbf{R}^3$. As was already noted, the mean curvature is determined by the way V^2 is embedded in $\mathbf{R}^3$, i.e. by both the first and the second fundamental forms. While studying the properties of the Gaussian curvature, we have found, in particular, two-dimensional surfaces of a given constant curvature. It was a rather simple matter to give several examples of surfaces of constant positive, zero, and negative curvature. Furthermore, it is easy enough to describe all surfaces V^2 of constant curvature, but we shall not prove this statement here. For example, a two-dimensional smooth, compact, closed Riemannian manifold of constant positive curvature is homeomorphic either to a sphere or to a projective plane. The description of surfaces of constant mean curvature is a much more complicated problem than in the case of the Gaussian curvature. Let us consider a surface of the mean zero curvature, the so-called minimal surface. Incidentally, such a surface is characterized by the property that its area is, locally, minimal in comparison with the area of other hypersurfaces that differ from the initial one only inside (any) ball of a rather small radius (see Fig. 4.35). The physical model for a minimal surface $V^2 \subset \mathbf{R}^3$ is a "soap film" formed on a wire loop when it is taken out of a vessel filled with soap water. Generally, the same loop can support several minimal films. Let us consider $V^2 \subset \mathbf{R}^3$ and derive an equation for a two-dimensional minimal surface. Since $H = \frac{GL - 2MF + EN}{EG - F^2}$, the equation $H = 0$ takes the form $GL -$

$2MF + EN = 0$. If the surface is given by the graph $z = f(x, y)$, then

$$ds^2 = (1 + f_x^2)\,dx^2 + 2f_x f_y\,dx\,dy + (1 + f_y^2)\,dy^2,$$

$$L = \frac{f_{xx}}{\sqrt{1+f_x^2+f_y^2}}, \quad M = \frac{f_{xy}}{\sqrt{1+f_x^2+f_y^2}}, \quad N = \frac{f_{yy}}{\sqrt{1+f_x^2+f_y^2}},$$

and hence

$$(1 + f_x^2)\,f_{xx} - 2f_{xy}f_x f_y + (1 + f_y^2)\,f_{yy} = 0.$$

The very form of this partial differential equation (whose solutions are minimal surfaces) shows that these solutions are rather complicated. (Example: a Euclidean plane $\mathbf{R}^2 \subset \mathbf{R}^3$ is a minimal surface, since $Q \equiv 0$ (verify!).)

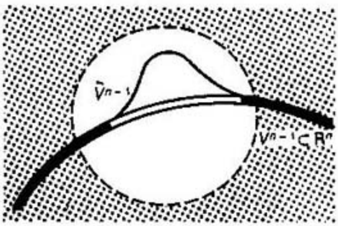

Figure 4.35. Volume of $\widetilde{V}^{n-1} \geqslant$ volume of V^{n-1}.

Here is a more complicated example of a non-compact, minimal surface. Let us consider in $\mathbf{R}^3$ two orthogonal straight lines l_1 and l_2 intersecting at point O. Fix the line l_1 and let the line l_2 move along l_1 at a constant velocity a and simultaneously rotate about l_1 at a constant angular speed ω (helical motion). The straight line l_2 sweeps out a two-dimensional smooth submanifold $V^2 \subset \mathbf{R}^3$ called a *right helicoid* (see Fig. 4.36).

Exercise: prove that a right helicoid is a minimal surface. To simplify calculations, we introduce on the helicoid the coordinates induced by the cylindrical coordinates in $\mathbf{R}^3$ with the axis $l_1 = Oz$.

Note that any minimal surface $V^2 \subset \mathbf{R}^3$ has a non-positive Gaussian curvature, since $\lambda_1 + \lambda_2 = 0$.

Another example deals with a contour $\Gamma \subset \mathbf{R}^3$, where by Γ we mean a smooth embedding of a set of non-intersecting circles in $\mathbf{R}^3$. Consider a non-compact minimal surface formed by the rotation of a smooth curve $\gamma(t)$, given by $y = a \cosh \frac{x}{a}$, about the x-axis. As

is known from mathematical analysis, this curve describes the shape of a freely sagging heavy chain fixed at two points A and B (Fig. 4.37).

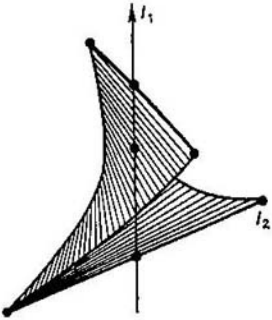

Figure 4.36

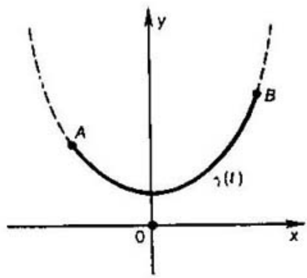

Figure 4.37

Here we assume that the chain is fixed so that the curve $y(x)$ does not intersect the x-axis. The gravity points downwards along the

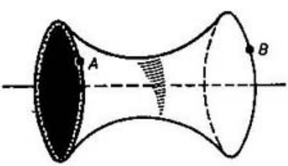

Figure 4.38

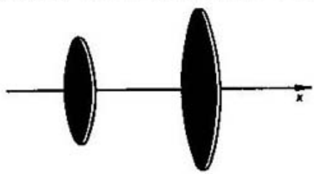

Figure 4.39

y-axis. The corresponding surface of rotation is called a *catenoid* (Fig. 4.38). Let us calculate H. It follows from Lemma 5 that

$$H = \lambda_1 + \lambda_2 = \frac{1}{y\sqrt{1+(y')^2}} - \frac{y''}{(1+(y')^2)^{3/2}}$$

$$= \left(a\cosh\frac{x}{a}\right)^{-1}\left(1+\sinh^2\frac{x}{a}\right)^{-\frac{1}{2}}$$

$$-\frac{1}{2}\cosh\frac{x}{a}\left(1+\sinh^2\frac{x}{a}\right)^{-\frac{3}{2}} = 0.$$

Thus, a catenoid is a minimal surface. The part of the catenoid which is bounded by the circles formed by the rotation of the points

A and B around the x-axis represents a minimal surface spanned by the contour Γ consisting of the two boundary circles. This example shows that the problem of finding a minimal surface with a given boundary contour does not have a unique solution. Indeed, besides the solution mentioned above (Fig. 4.38), there exists another minimal film with the same boundary contour, namely, two disks spanned by the boundary circles (Fig. 4.39). Whereas this minimal film exists

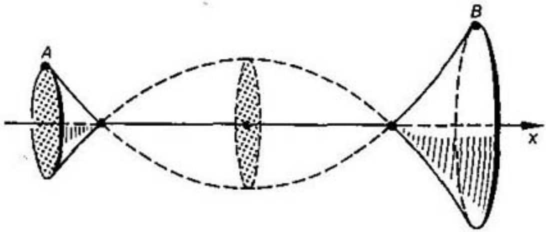

Figure 4.40

for any two arbitrary points A and B, a catenoid cannot always be spanned by the boundary circles. Apparently, if the distance between the points A and B is large enough, no catenoid can be constructed (Fig. 4.40). This becomes especially clear when we begin to draw the boundary circles apart, thereby extending the catenoid constructed for sufficiently close points A and B. The procedure is shown in Fig. 4.41. Upon extension the soap film breaks.

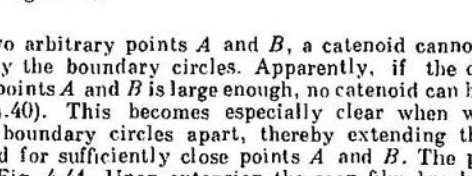

Figure 4.41

Another example of the contour that admits two solutions of the equation $H = 0$ is shown in Fig. 4.42. Here the two minimal films are homeomorphic to each other. The shape of the minimal films changes, depending on the way the circle S^1 is embedded in $\mathbf{R}^3$. For the standard embedding of S^1 in the plane (x, y) there is only one minimal film spanned by this contour, and this film coincides with the disk $x^2 + y^2 \leqslant 1$. If S^1 makes two revolutions about the z-axis, the solution of the equation $H = 0$ represents a Möbius band (see

Fig. 4.43). In the case of three revolutions, the solution is a "triple Möbius band" (Fig. 4.44). It can be obtained by moving a "trefoil", composed of three segments of equal length intersecting at the same angle $2\pi/3$, along a circle standardly embedded in a plane in such a way that after a complete revolution along the circle the trefoil

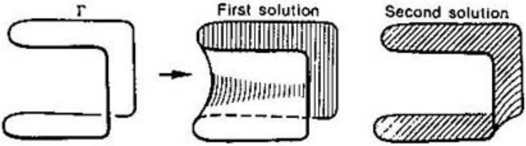

Figure 4.42

turns into itself with a rotation through $2\pi/3$ (see Fig. 4.44). The construction of such a surface is similar, in certain senses, to the construction of a right helicoid. A "triple Möbius band" is homeomorphic to the surface with self-intersections shown in Fig. 4.45.

Unlike the examples considered above, the minimal film shown in Fig. 4.44 has many singular points, i.e. the points any open neigh-

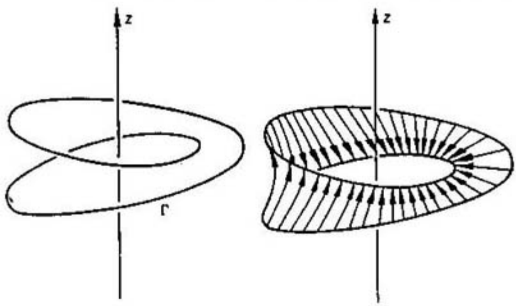

Figure 4.43

bourhood of which is not homeomorphic to a disk. The neighbourhood of each such point has the structure shown in Fig. 4.46, i.e. three half-disks glued along the common diameter. It is a simple matter to understand that any singular point of a two-dimensional minimal film, which is a non-compact surface without boundary, has the structure shown in Fig. 4.46. Indeed, if two half-disks converge at a singular point, then the neighbourhood of the point is homeomorphic to a disk. In the case of four half-disks convergent at a singular

point (see Fig. 4.47), there exists an area-reducing deformation of the neighbourhood. The total length of the segments AB, AC, AD,

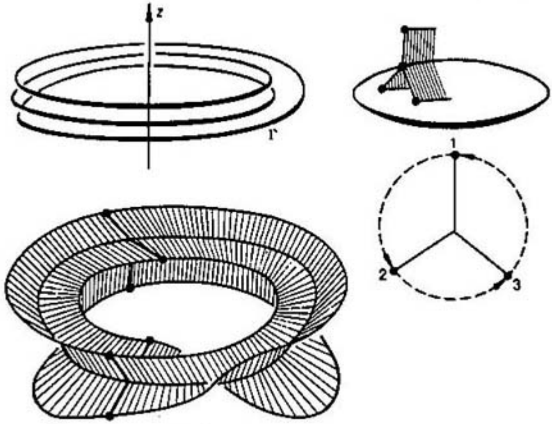

Figure 4.44. $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$.

and AR exceeds the total length of $A'B$, $A'R$, $A'A''$, $A''C$, and $A''D$ (the decomposition of a four-fold singularity into three-fold singu-

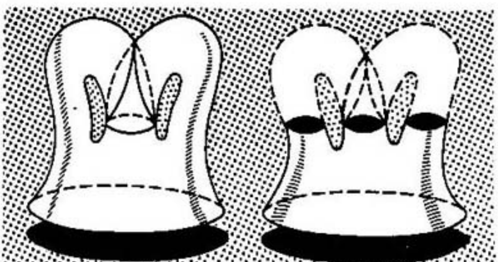

Figure 4.45

larities). For a minimal film with boundary this statement becomes invalid. An example is given in Fig. 4.48 where four-fold singular points fill the segment AB. Interestingly, the same contour Γ, but

made of a wire of a finite thickness, can span another soap film with three-fold singular points located along AB (see Fig. 4.49).

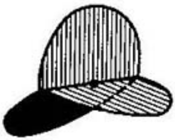

Figure 4.46

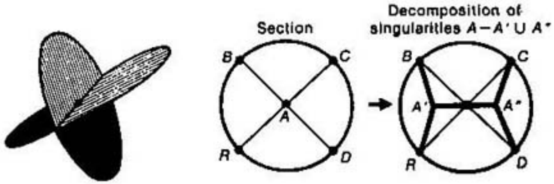

Figure 4.47

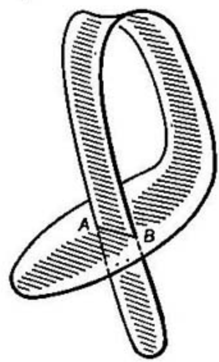

Figure 4.48

The so-called harmonic surfaces $V^2 \subset \mathbf{R}^3$ are closely related to minimal surfaces. Let V^2 be given as $\mathbf{r} = \mathbf{r}(u, v)$, where u and v are curvilinear coordinates on V^2.

Definition. The radius vector $\mathbf{r}(u, v)$ is called *harmonic* with respect to the coordinates u, v if $\frac{\partial^2 \mathbf{r}}{\partial u^2} + \frac{\partial^2 \mathbf{r}}{\partial v^2} = 0$, i.e. $\Delta \mathbf{r} = 0$, where Δ is the Laplacian in the coordinates u, v.

The radius vector $\mathbf{r}(u, v)$ which is harmonic in the coordinates (u, v) need not be necessarily harmonic in other coordinates u', v'.

Definition. The surface $V^2 \subset \mathbf{R}^3$ is called *harmonic* if it can be defined by a harmonic radius vector $\mathbf{r}(u, v)$ in curvilinear coordinates u, v.

The radius vector $\mathbf{r}(u, v)$ is said to be *minimal* if its mean curvature is identically zero. The function $H = \lambda_1 + \lambda_2$ is a scalar and, in

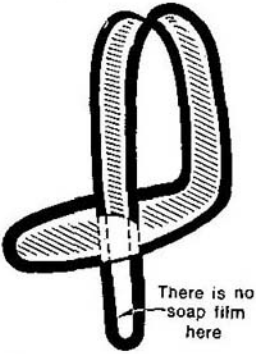

Figure 4.49

particular, remains unchanged under a regular coordinate transformation on a plane, so that if a radius vector is minimal relative to a coordinate system, it will also be minimal relative to any other regular coordinate system. Thus, the concept of a minimal film does not depend on the coordinates valid on this film. This is not however the case for a harmonic surface, and one usually speaks about a harmonic mapping $\mathbf{r}: D(u, v) \to \mathbf{R}^3(x, y, z)$, where $D(u, v)$ is a domain on the plane (u, v) and $\mathbf{r}(u, v)$ is the mapping defining the surface $V^2 \subset \mathbf{R}^3$. The mapping $\mathbf{r}$, which is harmonic in some coordinates, will not generally be harmonic in other coordinates (give an example!). Here is an example of a harmonic surface: let $\mathbf{r}(x, y)$ be defined by the formula $r(x, y) = (x, y, x^2 - y^2)$, where x, y, and z are Cartesian coordinates in $\mathbf{R}^3$, i.e. the surface V^2 is given by the graph $z = x^2 - y^2$ referred to a Cartesian coordinate system. Apparently, $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\mathbf{r} = 0$, i.e. the surface $z = x^2 - y^2$ is harmonic but it is not minimal because $H = 0$ only at the point $(0, 0)$, while at other points $H \neq 0$.

We recall that curvilinear coordinates (u, v) on a plane $V^2 \subset \mathbf{R}^3$ are called *conformal* if in these coordinates the metric $Edu^2 -$

$2Fdudv + Gdv^2$ induced on V^2 by the ambient Euclidean metric is diagonal, i.e. $E = G$ and $F = 0$.

Remark. Let us consider a two-dimensional smooth Riemannian manifold supplied with the Riemannian metric ds^2 with real-analytic coefficients E, F, and G which are considered as functions of the local curvilinear coordinates u, v (this metric is not necessarily induced by the embedding of M^2 in $\mathbf{R}^3$). Then for any point $P \in M^2$ there exists a neighbourhood $U = U(P)$ such that in U we can introduce coordinates p, q (they are real-valued analytic functions of the initial coordinates) in which ds^2 is of the form $\lambda(p, q)(dp^2 + dq^2)$, i.e. the coordinates p, q are conformal. The proof is not difficult, but requires the use of the theorem on the existence of a solution to a special partial differential equation (the Beltrami-Laplace equation), and this theorem is beyond the scope of our course.

What is a relationship between a harmonic and a minimal vector in $\mathbf{R}^3$? The example of a harmonic vector which is not minimal has already been given. Is the converse statement valid, i.e. is any minimal vector a harmonic one? The answer is no (give an example!). Yet, the following statement holds true.

Statement 2. *A minimal vector written in conformal coordinates is a harmonic one.*

Proof. Let (u, v) be conformal coordinates and let $\mathbf{r}(u, v)$ be the vector describing a minimal film. We recall that $E = G = \langle \mathbf{r}_v, \mathbf{r}_v \rangle = \langle \mathbf{r}_u, \mathbf{r}_u \rangle$, $F = \langle \mathbf{r}_u, \mathbf{r}_v \rangle$. We need to prove that $\mathbf{r}_{uu} + \mathbf{r}_{vv} = 0$. Put $\alpha = \mathbf{r}_{uu} + \mathbf{r}_{vv}$. Since $H = 0$, we have $GL - 2MF + EN = 0$, i.e. $L + N = 0$, whence $\langle \alpha, \mathbf{n} \rangle = \langle \mathbf{r}_{uu}, \mathbf{n} \rangle + \langle \mathbf{r}_{vv}, \mathbf{n} \rangle = L + N = 0$, where $\mathbf{n}$ is a normal to V^2. It remains to prove that $\langle \alpha, \mathbf{r}_u \rangle = \langle \alpha, \mathbf{r}_v \rangle = 0$, since in this case the scalar products of α and the orthogonal vectors $\mathbf{r}_u$, $\mathbf{r}_v$ will vanish, i.e. $\mathbf{n}$ is a zero vector. Differentiating the identities $E = G$ and $F = 0$, we obtain $\langle \mathbf{r}_{uu}, \mathbf{r}_u \rangle = \langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle$, $\langle \mathbf{r}_{uv}, \mathbf{r}_u \rangle = \langle \mathbf{r}_{vv}, \mathbf{r}_v \rangle$, $\langle \mathbf{r}_{uu}, \mathbf{r}_v \rangle = -\langle \mathbf{r}_u, \mathbf{r}_{uv} \rangle$, $\langle \mathbf{r}_{vu}, \mathbf{r}_v \rangle = -\langle \mathbf{r}_u, \mathbf{r}_{vv} \rangle$, whence $\langle \alpha, \mathbf{r}_u \rangle = \langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle - \langle \mathbf{r}_{vu}, \mathbf{r}_v \rangle = 0$. Similarly, $\langle \alpha, \mathbf{r}_v \rangle = 0$. The statement is proved.

Remark. We have asserted above (without proof) that conformal coordinates can be introduced in a neighbourhood of any point on a two-dimensional real-valued analytic surface. It can be proved that any minimal film can be defined by a real-valued analytic vector and, therefore, conformal coordinates can always be introduced in a neighbourhood of any point on a minimal surface (this fact is not trivial).

Problems

1. Prove that if the Gaussian and mean curvatures of a surface in $\mathbf{R}^3$ are identically zero, the surface is a plane.

2. Let the surface S be formed by tangent straight lines to a curve.

Express the Gaussian and mean curvatures of S in terms of the curvature and torsion of the curve.

3. Demonstrate in the preceding problem that the metric on the surface depends only on the curvature of the curve.

4.3. TRANSFORMATION GROUPS

4.3.1. SIMPLE EXAMPLES OF TRANSFORMATION GROUPS

In this section we shall discuss the basic examples of the transformation groups of metrics, i.e. such transformations of a manifold that preserve the metric. Consider a Riemannian manifold M^n with the metric g_{ij}.

Definition. The diffeomorphism f of a manifold M^n onto itself is a *length-preserving mapping* or an *isometry* if this mapping sends the Riemannian metric g_{ij} into itself, i.e. the following identity holds true:

$$g_{kp}(y) = g_{ij}(x(y)) \frac{\partial x^i(y)}{\partial y^k} \frac{\partial x^j(y)}{\partial y^p},$$

where $y^1, \ldots, y^n$ are local coordinates in a neighbourhood of a point $y \in M^n$, $x^1, \ldots, x^n$ are local coordinates in a neighbourhood of a point $x \in M^n$, and $x^i = x^i(y^1, \ldots, y^n)$, $1 \leqslant i \leqslant n$, are the functions defining (locally) the mapping f, with $x = f(y)$.

This is the "coordinate" definition of an isometry. It is sometimes convenient to use an invariant definition without any reference to local coordinates which cannot be chosen uniquely. Under a mapping f, the differential df sends T_yM^n onto T_xM^n, the latter mapping being a linear isomorphism since f is a diffeomorphism. In each of the tangent spaces T_yM^n and T_xM^n one can define scalar products $\langle , \rangle_y$ and $\langle , \rangle_x$, respectively, induced by the Riemannian metric, i.e. in the local coordinates $y^1, \ldots, y^n$, we have $\langle \mathbf{a}, \mathbf{b} \rangle_y = g_{ij}(y) a^i b^j$, where $\mathbf{a}, \mathbf{b} \in T_yM^n$.

Definition. The diffeomorphism f of a manifold M^n onto itself is called an *isometry* if $\langle \mathbf{a}, \mathbf{b} \rangle_y = \langle df(\mathbf{a}), df(\mathbf{b}) \rangle_x$ for any $\mathbf{a}, \mathbf{b} \in T_yM^n$; $x = f(y)$.

Lemma 1. *The coordinate and invariant definitions of an isometry are equivalent.*

Proof. Let $\mathbf{a} \in T_yM^n$, $\mathbf{a} = (a^1, \ldots, a^n)$ in the local coordinates $y^1, \ldots, y^n$. Hence, $df(\mathbf{a}) \in T_xM^n$ is of the form

$$(df(\mathbf{a}))^i = \frac{\partial x^i(y)}{\partial y^k} a^k,$$

since $df\colon T_yM^n \to T_xM^n$ is given by the Jacobi matrix. Thus,

$$\langle df(\mathbf{a}), df(\mathbf{b}) \rangle_x = g_{ij}(x(y)) \frac{\partial y^i(y)}{\partial y^k} \frac{\partial x^j(y)}{\partial y^p} a^k a^p = g_{kp} a^k a^p,$$

which proves the lemma.

Lemma 2. *The set of all isometries of a Riemannian manifold M^n forms a group (in the algebraic sense).*

Proof. That the composition of isometries is an isometry follows from the rule of differentiation of a composite function and from the law of transformation of g_{ij} under coordinate substitution. Also, that f^{-1} is an isometry stems from the fact that the Jacobi matrix of f^{-1} is the inverse of the Jacobi matrix $J(f)$. An identity transformation should be chosen as the unit element of the group. The lemma is proved.

The *isometry group* of a Riemannian manifold M^n is usually endowed with a topology in the mapping space and is denoted by Iso (M^n). Let us consider simple **examples.**

1. As M^1 we shall take the real axis (a non-compact one-dimensional manifold) with the Euclidean metric $ds^2 = dx^2$, where x is the coordinate on the axis. Let f be the diffeomorphism of $\mathbf{R}^1$ onto itself defined by a strictly increasing (or decreasing) function $x = f(y)$; the condition that f is an isometry implies that $ds^2 = (f'_y)^2 dy^2 = dy^2$, i.e. $(f'_y)^2 = 1$, so that f is either $f(y) = y + a$ or $f(y) = -y + b$, where a and b are arbitrary constants. Thus, the isometry group of a real axis is homeomorphic to a pair of real axes (proper isometries, which preserve the axis orientation, and improper isometries).

This example is virtually the unique one to find a complete isometry group of a manifold on an elementary level. The fact is that it is rather difficult to prove the completeness of a given subgroup in the isometry group. We succeeded in doing so only because the smooth function with a constant derivative is linear. Later we shall introduce the concept of a geodesic for a Riemannian metric and shall be able to prove the completeness of some subgroups in the isometry group.

2. Let us consider a Euclidean two-dimensional plane and find the isometry group that preserves point 0, the origin. We shall seek the isometries among linear transformations of the plane (it can be demonstrated that any isometry of a plane is linear, but here we shall not concentrate on this topic). The condition that the metric $dx^2 + dy^2$, $g_{ij} = \delta_{ij}$, is invariant can be written as the matrix equation $E = AA^{\mathrm{T}}$, where $A\colon \mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation. This is the definition of the orthogonal group, the group $O(2)$ in our case, which consists of the matrices $\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$ (proper rotations) and $\begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}$ (improper rotations or reflections). Proper rotations form a subgroup in $O(2)$ denoted by $SO(2)$; improper rotations do not form a subgroup. The subgroup $SO(2)$ is an invariant subgroup in $O(2)$; hence, there is valid the factor group

$O(2)/SO(2)$ isomorphic to $\mathbf{Z}_2$ (the second-order cyclic group). The group $O(2)$ is a subgroup in the isometry group of a circle endowed with the standard Riemannian metric $ds^2 = d\varphi^2$. Since $O(2)$ consists of matrices, this group becomes a topological space if the angle φ is associated with each matrix (for proper rotations). Thus, the set of the matrices that form $O(2)$ is homeomorphic to two copies of a circle; in other words, $O(2)$ can be provided with the structure of a smooth one-dimensional closed (disconnected) manifold (Fig. 4.50).

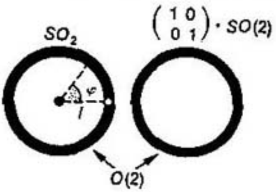

Figure 4.50

Prove that $O(2)$ coincides with the isometriy group of a circle with the metric $ds^2 = d\varphi^2$. *Hint:* use the procedure similar to that for the group Iso $(\mathbf{R}^1)$.

3. The motion of a Euclidean plane can be written as $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where $A \in O(2)$ and the vector $\mathbf{b}$ defines a translation on the plane. Apparently, all such transformations preserve the Euclidean metric (verify!). As is shown below, there are no isometries of a plane other than these transformations. This group can be represented in the matrix form

$$T = \left|\begin{array}{c|c} A & \mathbf{b} \\ \hline 0 \quad 0 & 1 \end{array}\right|.$$

(Prove that the coorrespondence $(\mathbf{y} = A\mathbf{x} + \mathbf{b}) \to T$ is an isomorphism of groups.) Hence, just as in the preceding example, the group Iso $(\mathbf{R}^2)$ can be transformed into a topological space homeomorphic to the direct product of a pair of circles and a Euclidean plane, and therefore this set can have the structure of a smooth three-dimensional manifold (this manifold is non-compact and disconnected and consists of two connected components).

4. Let us consider indefinite metrics. Define on $\mathbf{R}^2$ an indefinite metric $-dx^2 + dy^2$ which transforms a two-dimensional space into a pseudo-Euclidean plane $\mathbf{R}^2_1$. The matrix of the first fundamental form B is constant: $B = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. Hence, in order to find all linear

homogeneous transformations preserving this matrix, we have to solve the equation $B = ABA^T$ where $A\colon \mathbf{R}_1^2 \to \mathbf{R}_1^2$ is a linear mapping. Representing this mapping in the form $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we arrive at the system of equations for a, b, c and d: $a^2 - b^2 = 1$, $ac = bd$, $d^2 - c^2 = 1$. This system has the solution:

$$A = \begin{pmatrix} \pm \cosh \psi & \pm \sinh \psi \\ \pm \sinh \psi & \pm \cosh \psi \end{pmatrix},$$

or

$$A = \begin{pmatrix} \pm \dfrac{1}{\sqrt{1-\beta^2}} & \pm \dfrac{\beta}{\sqrt{1-\beta^2}} \\ \pm \dfrac{\beta}{\sqrt{1-\beta^2}} & \pm \dfrac{1}{\sqrt{1-\beta^2}} \end{pmatrix}, \quad \text{where} \quad \frac{b}{a} = \beta,\ \beta = \tanh \psi.$$

and the following sign combinations are possible:

$$\begin{pmatrix} + & + \\ + & + \end{pmatrix} \in \mathfrak{G}_1, \quad \begin{pmatrix} - & - \\ - & - \end{pmatrix} \in \mathfrak{G}_2, \quad \begin{pmatrix} + & - \\ + & - \end{pmatrix} \in \mathfrak{G}_3, \quad \begin{pmatrix} - & + \\ - & + \end{pmatrix} \in \mathfrak{G}_4.$$

This is a complete set of combinations. For example, $\begin{pmatrix} + & - \\ - & + \end{pmatrix}$ belongs to $\mathfrak{G}_1$, since substitution of $-\psi$ for ψ transforms $\begin{pmatrix} + & - \\ - & + \end{pmatrix}$ into $\begin{pmatrix} + & + \\ + & + \end{pmatrix}$ (we recall that $\sinh(-\psi) = -\sinh \psi$, $\cosh(-\psi) = \cosh \psi$). Thus, $\mathfrak{G} = \mathfrak{G}_1 \cup \mathfrak{G}_2 \cup \mathfrak{G}_3 \cup \mathfrak{G}_4$ (verify!) and $\mathfrak{G}_i \cap \mathfrak{G}_j = \varnothing$, provided $i \neq j$. Indeed, suppose, for example, that $\mathfrak{G}_1 \cap \mathfrak{G}_2 \neq \varnothing$, then $\cosh \varphi = -\cosh \psi$ and $\sinh \varphi = -\sinh \psi$, which is impossible since $\cosh \varphi > 0$ for any φ. It can be shown in a similar way that $\mathfrak{G}_i \cap \mathfrak{G}_j = \varnothing$.

Since $\mathfrak{G}$ is the group of homogeneous isometries of $\mathbf{R}_1^2$, transformations of this group map the set $\{-x^2 + y^2 = 1\} \cup \{-x^2 + y^2 = -1\}$, i.e. a pair of pseudo-circles of a real and an imaginary radius, into itself. In the case of a Euclidean plane $\mathbf{R}^2$ each rotation was described by the angle φ of the rotation of an orthogonal frame; an analogous parameter can be introduced for a pseudo-Euclidean plane $\mathbf{R}_1^2$. Let us consider an orthogonal frame $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$. Then the isometry A transforms this frame as is shown in Fig. 4.51. Also, instead of an ordinary Euclidean rotation angle φ we shall introduce the angle of hyperbolic rotation ψ by setting $\beta = \tanh \psi$ (see above); in this case $\mathfrak{G}$ becomes a group of hyper-

bolic rotations. Recall that the group of orthogonal transformations of a Euclidean plane consists of two connected components. The group of hyperbolic rotations has a more complicated structure: it consists of four connected components, namely the sets $\mathfrak{G}_1$, $\mathfrak{G}_2$, $\mathfrak{G}_3$, and $\mathfrak{G}_4$, i.e.

$$\left\{\begin{pmatrix} + & + \\ + & + \end{pmatrix}, \begin{pmatrix} - & - \\ - & - \end{pmatrix}, \begin{pmatrix} + & - \\ + & - \end{pmatrix}, \begin{pmatrix} - & + \\ - & + \end{pmatrix}\right\},$$

$-\infty < \psi < +\infty$. Since we have realized $\mathfrak{G}$ as a group of matrices, the group $\mathfrak{G}$ can be embedded as a subset in a four-dimensional real Euclidean space of all matrices of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $(a, b, c, d \in \mathbf{R})$ and therefore $\mathfrak{G}$ inherits the topology which turns

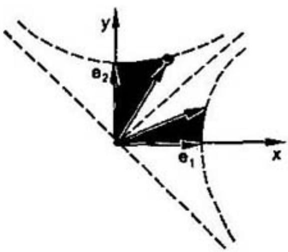

Figure 4.51

$\mathfrak{G}$ into a topological space. Each of the subsets $\mathfrak{G}_i$, $1 \leqslant i \leqslant 4$, is pathwise connected relative to this topology. Indeed, consider, for example, matrices of the type of $\mathfrak{G}_1$. Then for any two matrices $\begin{pmatrix} \cosh \psi_1 & \sinh \psi_1 \\ \sinh \psi_1 & \cosh \psi_1 \end{pmatrix}$ and $\begin{pmatrix} \cosh \psi_2 & \sinh \psi_2 \\ \sinh \psi_2 & \cosh \psi_2 \end{pmatrix}$ we can choose a continuous path $\gamma(t)$ which connects them in the set $\mathfrak{G}_1$: namely, $\begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} = \gamma(t)$, $\psi_1 \leqslant t \leqslant \psi_2$. The pathwise connectedness of the remaining subsets $\mathfrak{G}_2$, $\mathfrak{G}_3$, $\mathfrak{G}_4$ can be proved exactly in the same manner. Among these four connected components only $\mathfrak{G}_1$, i.e. $\begin{pmatrix} \cosh \psi & \sinh \psi \\ \sinh \psi & \cosh \psi \end{pmatrix}$, is a subgroup (verify!). None of the remaining components is a subgroup. For example, the product of two matrices

of the form $\begin{pmatrix} - & - \\ - & - \end{pmatrix}$ gives $\begin{pmatrix} + & + \\ + & + \end{pmatrix}$, i.e. if $\alpha, \beta \in \mathfrak{G}_2$, then $\alpha \cdot \beta \in \mathfrak{G}_1$ and $\alpha \cdot \beta \notin \mathfrak{G}_2$. The identity matrix belongs to $\mathfrak{G}_1$. The group $\mathfrak{G}$ is a subset in the four-dimensional space of the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and each $\mathfrak{G}_i$, $1 \leqslant i \leqslant 4$, is homeomorphic to the real axis. This homeomorphism (say, for $\mathfrak{G}_1$) is realized by associating with

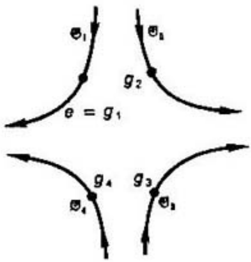

Figure 4.52. $g_1 := e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $g_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $g_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $g_4 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

each matrix $\begin{pmatrix} \cosh \psi & \sinh \psi \\ \sinh \psi & \cosh \psi \end{pmatrix}$ the angle ψ, the correspondence being one-to-one and continuous (see Fig. 4.52). The group $\mathfrak{G}$ maps the pseudo-circle of imaginary radius $-x^2 + y^2 = -1$ into itself. Figure 4.53 illustrates four transformations involving $\mathfrak{G}_1$, $\mathfrak{G}_2$, $\mathfrak{G}_3$, and $\mathfrak{G}_4$. The group $\mathfrak{G}$ transforms a pseudo-circle of real radius in a similar way.

The group $\mathfrak{G}$ is commutative (verify!). In this respect it resembles the group $O(2)$, which is also commutative. The subgroup $\mathfrak{G}_1$ is an invariant subgroup in $\mathfrak{G}$, since $g^{-1}qg$, where $q \in \mathfrak{G}_1$, $g \in \mathfrak{G}$, is, as before, transformation of type 1, for it is obtained from transformation of type 1 by hyperbolic rotation. Thus, the factor group $\mathfrak{G}/\mathfrak{G}_1$ is valid and its order is equal to the number of connected components in $\mathfrak{G}$, i.e. 4. Since $\mathfrak{G}$ is commutative, $\mathfrak{G}/\mathfrak{G}_1$ is also commutative. There exist only two commutative groups of the fourth order: namely, $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ and $\mathbf{Z}_4$. A question arises: to which group among these two is $\mathfrak{G}/\mathfrak{G}_1$ isomorphic? Let us compile the table for the multiplication of g_1, g_2, g_3, and g_4 which are representatives in their con-

nected components (see above). Calculation yields

	1	2	3	4
1	1	2	3	4
2	2	1	4	3
3	3	4	1	2
4	4	3	2	1

Thus, $\mathfrak{G}/\mathfrak{Z}_1$ is isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. This is a distinction between the rotation group for a pseudo-circle and the rotation group for an ordinary circle.

5. Let us consider the isometry group of a two-dimensional sphere defined as a Riemannian manifold with the metric induced by a standard embedding in $\mathbf{R}^3$. We first turn to the case of $\mathbf{R}^n$ and find the group of linear homogeneous transformations A preserving the Euclidean metric $ds^2 = \sum_{i=1}^{n} (dx^i)^2$. Since the matrix (g_{ij}) is of the form $E = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$, we have $E = AA^T$; the solutions of this equation are represented by orthogonal matrices which form the group $O(n)$. This group contains the subgroup $SO(n)$ consisting of proper rotations (with the determinant equal to $+1$); the remaining (improper) rotations do not form a subgroup. The subgroup $SO(n)$ is an invariant subgroup in $O(n)$ and the factor group $O(n)/SO(n)$ is isomorphic to $\mathbf{Z}_2$. Let $n = 3$, then $O(3)$ preserves the Euclidean metric in $\mathbf{R}^3$ and therefore maps the sphere S^2, $x^2 + y^2 + z^2 = R^2 = \text{const}$, into itself. Hence, $O(3)$ is a subgroup in $\mathrm{Iso}(S^2)$. Furthermore we shall demonstrate that $O(3) \equiv \mathrm{Iso}(S^2)$. Consider $SO(3) \subset O(3)$. Since $SO(3)$ is realized as a subset in the space of all 3×3 matrices with real coefficients, identified with $\mathbf{R}^9$, this subgroup is provided with an induced topology and becomes a topological space.

Lemma 3. *The group $SO(3)$, as a topological space, is homeomorphic to a three-dimensional projective space $\mathbf{RP}^3$.*

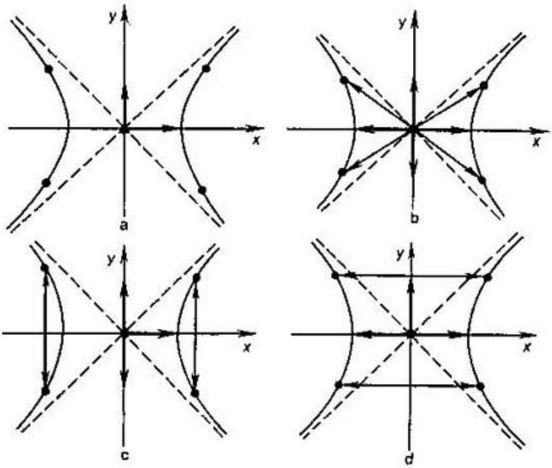

Figure 4.53. (a) $g_1 := e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, identity mapping; (b) $g_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, reflection relative to the origin; (c) $g_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, reflection relative to the x-axis; (d) $g_4 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, reflection relative to the y-axis.

Proof. Let A be an element of $SO(3)$; then there exists in $\mathbf{R}^3$ a fixed axis $l(A)$ such that the action of A in $\mathbf{R}^3$ is rotation about $l(A)$ through an angle $\varphi(A)$. If $A \neq E$, then $l(A)$ is defined uniquely. Consider a plane $\Pi(A)$ through the origin O perpendicular to the axis $l(A)$; choose in $\Pi(A)$ an arbitrary vector $\mathbf{e}_1$ and let $\mathbf{e}_2$ be the vector obtained from $\mathbf{e}_1$ by the rotation through the angle $\varphi(A)$ (see Fig. 4.54). Complete $\mathbf{e}_1$, $\mathbf{e}_2$ to a coordinate frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, so that its orientation coincides with the orientation of a fixed frame $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ in $\mathbf{R}^3$. Then $l(A)$ becomes a real axis if we can define on it direction with the aid of $\mathbf{e}_3$ and reckon the value of $\varphi(A)$. Thus, we have uniquely associated with each element $A \in SO(3)$ a point in $\mathbf{R}^3$; denote this point by $P(l, \varphi)$. Apparently, $P(l, \pi) = P(l, -\pi)$ since rotations about $l(A)$ through π and $-\pi$ coincide. If $|\varphi(A)| < \pi$, $P(l, \varphi)$ corresponds only to the rotation of A. By continuously

varying A, we continuously change $P(l, \varphi)$, the converse is also true. Thus, we have realized a one-to-one continuous correspondence between orthogonal transformation of A and points of a three-dimensional ball of radius π whose antipodal points $P(l, \pi)$ and

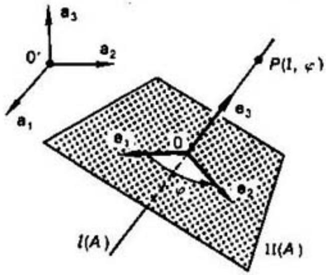

Figure 4.54

$P(l, -\pi)$ are glued on the boundary (i.e. on the sphere of radius π). It remains to prove that this ball glued on boundaries is homeomorphic to $\mathbf{RP}^3$. According to one of the definitions of $\mathbf{RP}^3$, this space is a sheaf of straight lines in $\mathbf{R}^4$ through the origin O; such

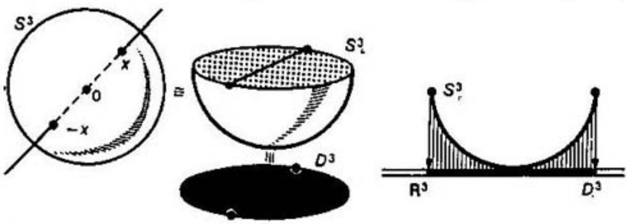

Figure 4.55

a model is equivalent to the one in which we choose S^3 and identify its antipodal points. To this end, we take a hemisphere S^3_+ and identify the antipodal points on its boundary, i.e. on the equator (Fig. 4.55). A hemisphere is diffeomorphic to a three-dimensional disk, and the diffeomorphism can be obtained by orthogonally projecting S^3_+ onto D^3 (see Fig. 4.55). Thus, $\mathbf{RP}^3$ is homeomorphic to D^3 with identified antipodal boundary points. The lemma is proved.

Since $\mathbf{RP}^3$ is pathwise connected, $O(3)$ consists of two path components.

6. Let us consider the isometry group of a Lobachevskian plane with a standard Riemannian metric. The Lobachevskian plane is assumed to be defined on the upper half-plane with the metric $\frac{dz\,d\bar{z}}{(z-\bar{z})^2}$. The isometry group of this metric will be sought among homographic transformations of the complex plane $\frac{az+b}{cz+d}$, $a, b, c, d \in \mathbf{C}$. Let us study how the Euclidean metric $dz\,d\bar{z}$ behaves under the transformation $w=\frac{az+b}{cz+d}$. Since $dw=\frac{ad-bc}{(cz+d)^2}\cdot dz$ (verify!), we have

$$dw\,d\bar{w}=\frac{|ad-bc|^2}{|cz+d|^4}\,dz\,d\bar{z}.$$

Thus, the metric is multiplied by a scalar variable factor, that is, homographic transformations are conformal; they preserve the cosines of the angles between intersecting curves. It remains to prove

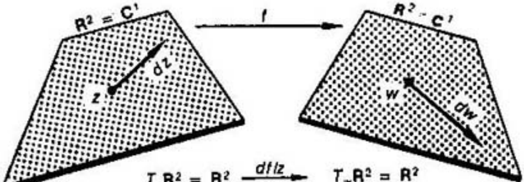

Figure 4.56

that the oriented angles are also preserved. Calculating the Jacobian $J\left(w=\frac{az+b}{cz+d}\right)$, we find that it is positive. Let $z \in \mathbf{R}^2$, $w \in \mathbf{R}^2$, $w=\frac{az+b}{cz+d}$, $dw=\lambda(z)\,dz$, $\lambda(z)=\frac{ad-bc}{(cz+d)^2}$, $T_z\mathbf{R}^2=\mathbf{R}^2$; $T_w\mathbf{R}^2=\mathbf{R}^2$ (see Fig. 4.56). Hence, the operation df of the mapping $f\colon z \to w$ is written as $df(z)=\lambda\cdot z$, where $\lambda \in \mathbf{C}$. Let $\lambda=u+iv$, $u, v \in \mathbf{R}$, then the Jacobi matrix in the real form is $\begin{pmatrix} u & v \\ -v & u \end{pmatrix}$, i.e. the Jacobian is equal to $u^2+v^2>0$. Out of all homographic transformations we choose those that map the upper half-plane into itself.

Lemma 4. *The transformation* $w=\frac{az+b}{cz+d}$ *maps the upper half-plane into itself if and only if* $(a, b, c, d)=\rho\,(a', b', c', d')$, *where* $a', b', c', d' \in \mathbf{R}$, $\rho \in \mathbf{C}$, $\rho \neq 0$; *i.e. all the coefficients* (a, b, c, d)

are proportional to the quadruple of real numbers (a', b', c', d') *and* $ad - bc > 0$.

Proof. Let the quadruple of the coefficients be proportional to the real quadruple (denoted by a, b, c, d). Apparently, the real axis is mapped into itself. We now demonstrate that if the point z belongs to the upper half-plane, its image $w = \frac{az+b}{cz+d}$ also belongs to this half-plane, i.e. $\operatorname{Im}\left(\frac{az+b}{cz+d}\right) > 0$. Indeed,

$$w = \frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} = \frac{ac(z\bar{z})+bd}{|cz+d|^2} + \frac{adz+bc\bar{z}}{|cz+d|^2},$$

$$\operatorname{Im} w = \frac{ad-bc}{|cz+d|^2}\cdot \operatorname{Im} z > 0,$$

since $ad - bc > 0$; $\operatorname{Im} z > 0$. Conversely, let $w = \frac{az+b}{cz+d}$ map the upper half-plane into itself. We have to prove that there exists a common factor ρ such that (a, b, c, d) are proportional to the real quadruple (a', b', c', d'). If $\bar{z} = z$, then $\bar{w} = w$, and for an arbitrary $x \in \mathbf{R}$ we obtain

$$\frac{ax+b}{cx+d} = \frac{\bar{a}x+\bar{b}}{\bar{c}x+\bar{d}}.$$

For $x = 0$, $\frac{b}{d} = \frac{\bar{b}}{\bar{d}} = \lambda$, $\lambda \in \mathbf{R}$. For $x \to \infty$,

$$\frac{a}{c} = \frac{\bar{a}}{\bar{c}} = \mu \in \mathbf{R},\ b = \lambda d,\ a = \mu c,$$

for $x = 1$,

$$\frac{a+b}{c+d} = \frac{\bar{a}+\bar{b}}{\bar{c}+\bar{d}} = \rho \in \mathbf{R},\ \mu c + \lambda d = \rho c + \rho d,\ (\mu - \rho)c = (\rho - \lambda)d.$$

In the case of general position, i.e. when $\mu - \rho \neq 0$, $\rho - \lambda \neq 0$ we obtain $c = \xi d$, $\xi \in \mathbf{R}$. Thus, all four complex numbers a, b, c, and d lie on a single straight line (Fig. 4.57). Multiplying these numbers by a complex factor, we can rotate this straight line so that it coincides with the real axis. The lemma is proved.

Lemma 5. *Any transformation* $w = \frac{az+b}{cz+d}$ *such that* $a, b, c, d \in \mathbf{R}$ *and* $ad - bc > 0$ *is an isometry of the Lobachevskian plane.*

Proof. We have

$$dw = \frac{ad-bc}{(cz+d)^2}\,dz, \quad \frac{dw\,d\bar{w}}{(w-\bar{w})^2} = \frac{dz\,d\bar{z}}{(z-\bar{z})^2},$$

which is what was required.

Since $ad - bc > 0$, we may take $ad - bc = 1$.

Proposition 1. *The isometry group of the Lobachevskian plane* $\mathrm{Iso}(L_2)$ *contains a subgroup isomorphic to the group* $SL(2, \mathbf{R})/\mathbf{Z}_2$, *i.e. to the factor group* $SL(2, \mathbf{R})$ *of* 2×2 *matrices with real coefficients*

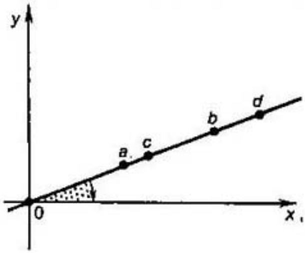

Figure 4.57

and determinant $+1$ *with respect to the subgroup* $\mathbf{Z}_2$ *consisting of the transformations* E *and* $-E$.

Proof. Let us consider $w = \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbf{R}$ and $ad - bc = 1$. According to Lemma 5, these transformations are isometries. The set of these isometries forms a group. Indeed,

$$\left(a'\left(\frac{az+b}{cz+d}\right)+b'\right)\Big/\left(c'\left(\frac{az+b}{cz+d}\right)+d'\right)=\frac{(a'a+cb')z+(a'b+db')}{(c'a+cd')z+(c'b+dd')}$$

is again a transformation with real coefficients and determinant $+1$. Since $ad - bc \neq 0$, there exists an inverse transformation of the same type. Consider the matrix group $SL(2, \mathbf{R})$, i.e. the group $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $a, b, c, d \in \mathbf{R}$ and $ad - bc = 1$, and construct the mapping φ: $SL(2, \mathbf{R}) \to \mathfrak{G}_1$, where $\mathfrak{G}_1$ is the group of transformations $w = \frac{az+b}{cz+d}$, $a, b, c, d \in \mathbf{R}$, $ad - bc = 1$. Put $\varphi\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{az+b}{cz+d}$. Obviously, φ is a homomorphism (verify!) and, furthermore, φ is an epimorphism. Find $\ker\varphi$. Clearly, $\varphi(g) = \varphi(-g)$, so that $\ker\varphi \supset \mathbf{Z}_2 = (E, -E)$. Prove that $\ker\varphi = \mathbf{Z}_2$. Let $\varphi(g) = \varphi(g')$, i.e. $\frac{az+b}{cz+d} = \frac{a'z+b'}{c'z+d'}$, whence

$$\frac{b}{b'} = \frac{d}{d'} = \lambda, \quad \frac{a}{a'} = \frac{c}{c'} = \mu, \quad \frac{a+b}{c+d} = \frac{a'+b'}{c'+d'}, \quad b = \lambda b', \; d = \lambda d',$$

$$a = \mu a', \; c = \mu c', \; \mu a'd' + \lambda b'c' = \mu c'b' + \lambda d'a',$$

$$(\mu - \lambda)(a'd' - b'c') = 0, \; \mu = \lambda, \; a'd' - b'c' = 1.$$

Thus, $\frac{b}{b'}=\frac{d}{d'}=\frac{a}{a'}=\frac{c}{c'}=\lambda$, i.e. $g'=\lambda g$, $\lambda=\pm 1$, $g'=\pm g$, which is what was required.

Problem. Prove that $SL(2, \mathbf{R})/\mathbf{Z}_2$ is a connected topological space.

One should not believe that only transformations $SL(2, \mathbf{R})/\mathbf{Z}_2$ are contained in the group $\mathrm{Iso}(L_2)$. Indeed, the transformation g_0: $z \to -\bar{z}$ maps the upper half-plane into itself and preserves the

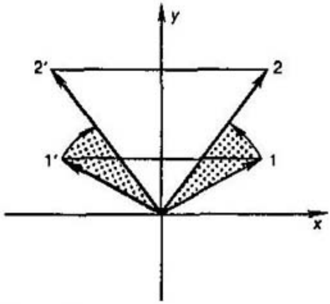

Figure 4.58

Lobachevskian metric (reflection about the y-axis). At the same time, g_0 is not of the form $\frac{az+b}{cz+d}$. Indeed, $\frac{az+b}{cz+d}$ are conformal (see above), i.e. they preserve oriented angles, while the transformation g_0: $z \to -\bar{z}$ is not conformal (see Fig. 4.58).

Thus we should also consider transformations $g(z)$ of the form $w=g(z)=-\frac{a\bar{z}+b}{c\bar{z}+d}$, where $a, b, c, d \in \mathbf{R}$, $ad-bc=1$. In other words, $w=\frac{a\bar{z}+b}{c\bar{z}+d}$, where $a, b, c, d \in \mathbf{R}$, $ad-bc=-1$. Let $\mathfrak{G}_2$ denote the set of all such transformations. The sets $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are homeomorphic, since any $g \in \mathfrak{G}_2$ is of the form $g=g_0 f$, where $f \in \mathfrak{G}_1$, and since g_0 and f are isometries, $g_0 f$ is an isometry as well. The homeomorphism is established by multiplying the set $\mathfrak{G}_1$ by $g_0 \in \mathfrak{G}_2$. Also $\mathfrak{G}_1 \cap \mathfrak{G}_2 = \varnothing$, since $\frac{az+b}{cz+d} \not\equiv \frac{\alpha\bar{z}+\beta}{\gamma\bar{z}+\delta}$. Indeed, $\frac{az+b}{cz+d}$ preserves oriented angles, while $\frac{\alpha\bar{z}+\beta}{\gamma\bar{z}+\delta}$ does not.

Lemma 6. *The set $\mathfrak{G}=\mathfrak{G}_1 \cup \mathfrak{G}_2=\{f\} \cup \{g_0 f\}$ is a group in which $\mathfrak{G}_1$ is a subgroup and $\mathfrak{G}_2$ is not a subgroup.*

Proof. Consider the set of all real-valued 2×2 matrices with the determinant ± 1. Apparently, this is a group; we denote it by $L(2, \mathbf{R})$, i.e. $\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},\ ad-bc=\pm 1\right\}$. This group, as a topological space, is disconnected, for it has two connected components: $L(2, \mathbf{R}) = L_1 \cup L_2$, where

$$L_1 = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},\ ad-bc=1\right\},\quad L_2 = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},\ ad-bc=-1\right\}.$$

The subgroup L_1 was denoted above by $SL(1, \mathbf{R})$. Let us construct $\varphi: L \to \mathfrak{G}$. If

$$A = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in L_1, \text{ then } \varphi(A) = f \in \mathfrak{G}_1,\quad f(z) = \frac{az+b}{cz+d},\quad ad-bc=+1,$$

if

$$B = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in L_2, \text{ then } \varphi(B) = g \in \mathfrak{G}_2,\ g(z) = \frac{a\bar{z}+b}{c\bar{z}+d},\quad ad-bc=-1.$$

Clearly, φ is a homomorphism (verify!). Furthermore, the mapping φ is an epimorphism, it is not one-to-one and has a kernel. To find the kernel, it is sufficient to find the inverse image of the identity element of the group $\mathfrak{G}$. As in Proposition 1, it is a simple matter to prove that $\ker(\varphi) = \mathbf{Z}_2$, where $\mathbf{Z}_2 = (+E, -E)$; this subgroup is the centre in $L(2, \mathbf{R})$. The lemma is proved. This means that we have also proved the following statement.

Lemma 7. *The group* $\mathfrak{G} = \mathfrak{G}_1 \cup \mathfrak{G}_2 = \{f\} \cup \{g\}$ *is isomorphic to* $L(2, \mathbf{R})/\mathbf{Z}_2$, *where*

$$L(2, \mathbf{R}) = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},\ ad-bc=\pm 1\right\},\quad \mathbf{Z}_2 = (E, -E).$$

Thus, in the isometry group of the Lobachevskian plane we have found a subgroup consisting of two path components, namely, $\mathfrak{G} \simeq L(2, \mathbf{R})/\mathbf{Z}_2$. How many parameters are needed to describe the elements of this group? Since $L(2, \mathbf{R}) = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},\ ad-bc=1\right\}$, the element $g \in \mathfrak{G}$ is determined by three independent parameters.

Lemma 8. *The group $SL(2, \mathbf{R})$, as a topological space, is homeomorphic to the direct product of a circle and a Euclidean plane and can, therefore, be provided with the structure of a smooth three-dimensional (non-compact) manifold. Correspondingly, $L(2, \mathbf{R})$ is homeomorphic to the direct product of two copies of a circle and a Euclidean plane.*

Proof. It is known from algebra that any linear homogeneous transformation of a Euclidean plane with the determinant $+1$ can

uniquely be represented as the composition of the proper rotation and triangular transformation (the basis orthogonalization theorem). Thus, any matrix $g \in SL(2, \mathbf{R})$ admits uniquely a representation as the matrix product $g = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} \alpha & \beta \\ 0 & \frac{1}{\alpha} \end{pmatrix}$. Since the matrices $\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$ form a circle and the matrices $\begin{pmatrix} \alpha & \beta \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ form a Euclidean plane, the lemma is proved.

The topological structure of $L(2, \mathbf{R})/\mathbf{Z}_2$ is more complicated. As is shown below, the subgroup $L(2, \mathbf{R})/\mathbf{Z}_2$ is, in fact, the whole group Iso (L_2). Note also that the isometry groups of a sphere and a Lobachevskian plane have the same dimension and can be endowed with the structure of a smooth three-dimensional manifold.

Remark on the isometry group of a pseudo-Euclidean space $\mathbf{R}_1^3$. We recall that the isometry group of a two-dimensional sphere coincides with the isometry group of $\mathbf{R}^3$ that preserves the point O fixed,

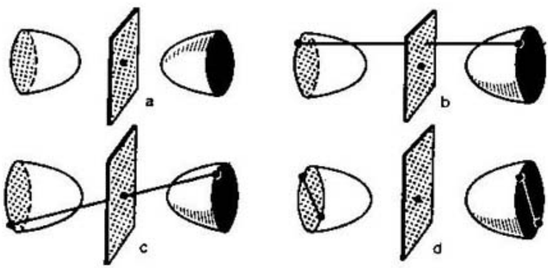

Figure 4.59. (a) $g_1 = e$, identity transformation; (b) g_2, reflection in the plane yz; (c) g_3, reflection at the point 0; (d) g_4, reflection in the plane yx.

i.e. with the group $O(3)$. Similarly, a Lobachevskian plane can be realized in $\mathbf{R}_1^3$ as one of the sheets of a two-sheet hyperboloid (a pseudosphere of imaginary radius), and any motion of the Lobachevskian plane will be induced by an isometry of $\mathbf{R}_1^3$. Since a pseudosphere of imaginary radius consists of two connected components, Iso L_2 is not the whole group Iso $(\mathbf{R}_1^3)_0$, where Iso $(\mathbf{R}_1^3)_0$ denotes all isometries of $\mathbf{R}_1^3$ that leave the point O fixed; reflection-induced trans-

formations, which represent two sheets of a hyperboloid, should also be included. Thus, the group $\mathrm{Iso}\,(\mathbf{R}_1^3)_0$ consists of four connected components. Figure 4.59 shows four transformations, g_1, g_2, g_3, and g_4, which preserve the pseudosphere $S_1^2 \supset L_2$ and belong to different connected components of the group $\mathrm{Iso}\,(\mathbf{R}_1^3)_0$.

7. Let us consider again the group of motions of a Euclidean plane. The transformations $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, $A \in O\,(2)$, $\mathbf{b} \in \mathbf{R}^2$, found above can be written in the complex form $w = az + b$, where $b \in \mathbf{C}$, $a \in \mathbf{C}$, $|a| = 1$, i.e. $w = e^{i\varphi}\cdot z + b$. This transformation group is isomorphic to the matrix group consisting of matrices $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$, where $a = e^{i\varphi}$. That the Euclidean metric is preserved directly follows from the identity $dw = adz$; $dwd\bar{w} = |a|^2\, dzd\bar{z}$.

Just like in the case of a plane, it is a simple matter to verify that the group $\mathrm{Iso}\,(\mathbf{R}^n)$ of linear isometries of a Euclidean space $\mathbf{R}^n$ can be represented as the transformation group $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where the matrix A belongs to the orthogonal group $O\,(n)$ and the vector $\mathbf{b}$ defines a translation. The group $\mathrm{Iso}\,(\mathbf{R}^n)$ is isomorphic to the matrix group consisting of matrices of the form

$$\left(\begin{array}{c|c} A & \mathbf{b} \\ \hline 0\ \ldots\ldots\ 0 & 1 \end{array}\right).$$

Thus, the group $\mathrm{Iso}\,(\mathbf{R}^n)$, as a topological space, is homeomorphic to the direct product $O\,(n) \times \mathbf{R}^n$, where $O\,(n)$ and $\mathbf{R}^n$ are considered as topological spaces. This decomposition is not, however, a group one.

4.3.2. MATRIX TRANSFORMATION GROUPS

We have seen that all the examples of groups considered above are topological spaces on which the structure of a smooth manifold can be introduced in a natural way. We have used the topology which arises on a transformation group after it is embedded in a matrix group whose topology is defined in a usual manner: matrices are assumed to be close if their elements are close. Thus, we arrive at a class of smooth manifolds such that points on them can be "multiplied", this multiplication satisfying all axioms of an algebraic group. Let us give the following definition.

Definition. A *Lie group* is a smooth manifold M^n endowed with two smooth mappings $f\colon M^n \times M^n \to M^n$ (multiplication) and $\nu\colon M^n \to M^n$ (construction of the inverse element), usually denoted as $f(x, y) = x \cdot y$, $\nu(x) = x^{-1}$, and having a marked point $e \in M^n$ which satisfies together with f and ν the relations: (1) $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, (2) $e \cdot x = x \cdot e = x$, $x \cdot x^{-1} = x^{-1} \cdot x = e$.

The operations f and ν are usually required to be continuous, but in all examples considered below the group operations are smooth, that is why smooth mappings f and ν have been used in the definition of the Lie group.

Definition. The set of elements g of a Lie group $\mathfrak{G}$ that can be connected by a continuous path with the unit element of $\mathfrak{G}$ is called the *connected component of the unit element of the group* $\mathfrak{G}$ and is denoted by $\mathfrak{G}_0$.

Statement 1. *The set $\mathfrak{G}_0$ is a subgroup in $\mathfrak{G}$. Furthermore, $\mathfrak{G}_0$ is an invariant subgroup in $\mathfrak{G}$, so there exists the factor group $\mathfrak{G}/\mathfrak{G}_0$.*

Proof. Let $g_1, g_2 \in \mathfrak{G}_0$; we need to prove that $g_1 \cdot g_2 \in \mathfrak{G}_0$. According to the definition of $\mathfrak{G}_0$, there exist continuous paths $\gamma_1(t)$ and $\gamma_2(t)$ such that $\gamma_1(0) = e$, $\gamma_1(1) = g_1$ and $\gamma_2(0) = e$, $\gamma_2(1) = g_2$. Consider the path $\gamma\colon I \to \mathfrak{G}$, where $\gamma(t) = \gamma_1(t) \cdot \gamma_2(t)$. Since multiplication is continuous in $\mathfrak{G}$, the path $\gamma(t)$ is continuous and therefore $g_1 \cdot g_2 \in \mathfrak{G}_0$ because $\gamma(0) = \gamma_1(0) \cdot \gamma_2(0) = e$; $\gamma(1) = \gamma_1(1) \cdot \gamma_2(1) = g_1 \cdot g_2$. Demonstrate that for any $g_0 \in \mathfrak{G}_0$ and any $g \in \mathfrak{G}$ the element gg_0g^{-1} belongs to $\mathfrak{G}_0$. Since $g_0 \in \mathfrak{G}_0$, there exists a continuous path $\gamma(t)$ such that $\gamma(0) = e$, $\gamma(1) = g_0$. Consider another path $\varphi(t) = g\gamma(t)g^{-1}$, it is continuous because the multiplication operation is continuous. At the same time, $\varphi(0) = e$, $\varphi(1) = gg_0g^{-1}$, i.e. $gg_0g^{-1} \in \mathfrak{G}_0$. The statement is proved.

We have already got acquainted with the examples of Lie groups. For instance, the group of orthogonal matrices is a Lie group (verify!). Let us consider the most typical examples of matrix groups. All these groups are Lie groups; we shall not, however, prove this fact, confining ourselves to certain particular cases.

1. **Full linear groups $\mathfrak{G}L(n, \mathbf{R})$ and $\mathfrak{G}L(n, \mathbf{C})$.** Let us consider a Euclidean space $\mathbf{R}^n$ and the set of all non-degenerate linear homogeneous transformations of $\mathbf{R}^n$ into itself, i.e. the set of all nonsingular $n \times n$ matrices with real coefficients. This set is denoted by $\mathfrak{G}L(n, \mathbf{R})$. The set $\mathfrak{G}L(n, \mathbf{C})$ is defined in a similar way.

Lemma 9. *The sets $\mathfrak{G}L(n, \mathbf{R})$ and $\mathfrak{G}L(n, \mathbf{C})$ are Lie groups.*

Proof. Let us consider, for certainty, the set $\mathfrak{G}L(n, \mathbf{R})$. That $\mathfrak{G}L(n, \mathbf{R})$ forms a group (in the algebraic sense) is obvious. It remains to prove that $\mathfrak{G}L(n, \mathbf{R})$ can be provided with the structure of a smooth manifold such that all group operations are smooth. Clearly, $\mathfrak{G}L(n, \mathbf{R}) = \mathbf{R}^{n^2} \setminus \{\det g = 0\}$, where the Euclidean space $\mathbf{R}^{n^2}$ is identified with the space of all matrices of order n (over the field $\mathbf{R}$). Since the equation $\det g = 0$ is polynomial, the

set $\mathbf{R}^{n^2} \setminus \{\det g = 0\}$ is open in $\mathbf{R}^{n^2}$, i.e. this set is a domain in $\mathbf{R}^{n^2}$ and, therefore, a smooth manifold of dimension n^2. Multiplication of matrices is a smooth operation because each element of the matrix AB is a second-order polynomial in the elements of the matrices A and B. Each element of the inverse matrix A^{-1} is a rational function of the elements of the marix A (the denominator of the function is non-zero since A is non-singular). The matrix E is the unity element of the group $\mathfrak{G}L(n, \mathbf{R})$. In a similar fashion we can demonstrate that $\mathfrak{G}L(n, \mathbf{C})$ is a Lie group. The lemma is proved.

2. **Special linear groups** $SL(n, \mathbf{R})$ **and** $SL(n, \mathbf{C})$. The group $SL(n, \mathbf{R})$ is defined as a subset in $\mathfrak{G}L(n, \mathbf{R})$ given by the equation $\det g = 1$. Clearly, this set is a group and a topologial space. The group $SL(n, \mathbf{R})$ is, in fact, a smooth manifold, but we shall not prove here this assertion. The group $SL(n, \mathbf{C})$ is defined as a subgroup in $\mathfrak{G}L(n, \mathbf{C})$ which satisfies $\det g = 1$. The dimension of $SL(n, \mathbf{R})$ is $n^2 - 1$ and that of $SL(n, \mathbf{C})$ is $2n^2 - 2$.

3. **Orthogonal groups** $O(n, \mathbf{R})$ **and** $O(n, \mathbf{C})$. Let us consider $\mathbf{R}^n$ with the bilinear form $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{n} a^i b^i$ which defines a Euclidean scalar product. The group $O(n, \mathbf{R})$ is defined as the group of real matrices A of order n that preserve this scalar product, i.e. satisfy the relation $\langle A\mathbf{a}, A\mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle$ for any $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$. The group $O(n, \mathbf{R})$ is usually denoted as $O(n)$. The group $O(n, \mathbf{C})$ is defined in a similar way. The group $O(n)$ contains a subgroup denoted by $SO(n)$ and called special orthogonal group: $g \in SO(n)$ if $\det g = 1$.

Lemma 10. *The group $SO(n)$ is pathwise connected and coincides with the unity component of $O(n)$. The factor group $O(n)/SO(n)$ is isomorphic to $\mathbf{Z}_2$, i.e. $O(n)$ consists of two connected components.*

Proof. It is known from algebra that for any element $g_0 \in SO(n)$ there exists an orthogonal transformation $g \in O(n)$ such that $a = g g_0 g^{-1}$ is matrix of the form

$$\left(\begin{array}{cc|ccc|cc}
\cos\varphi_1 & \sin\varphi_1 & & & & & \\
-\sin\varphi_1 & \cos\varphi_1 & & & 0 & & \\
\hline
& & \ddots & & & & \\
& & & & \ddots & & \\
\hline
& & & 0 & & \cos\varphi_k & \cos\varphi_k \\
& & & & & -\sin\varphi_k & \cos\varphi_k
\end{array}\right)$$

(provided $n = 2k$) and of the form

$$
\left|\begin{array}{cc|c|cc|c}
\cos\varphi_1 & \sin\varphi_1 & & & & \\
-\sin\varphi_1 & \cos\varphi_1 & & & 0 & \\
\hline
& & \ddots & & & \\
\hline
& & & \cos\varphi_k & \sin\varphi_k & 0 \\
& 0 & & -\sin\varphi_k & \cos\varphi_k & \\
\hline
& & & & 0 & 1
\end{array}\right|
$$

(provided $n = 2k + 1$). As a continuous path $\gamma(t)$ connecting g_0 with the identity matrix E we may consider the following family of matrices:

$$
a(t) = \left|\begin{array}{cc|c}
\cos\varphi_1 t & \sin\varphi_1 t & 0 \\
-\sin\varphi_1 t & \cos\varphi_1 t & \\
\hline
& 0 & \ddots
\end{array}\right|, \quad \text{where } \gamma(t) = g^{-1} a(t)\, g.
$$

Thus, $SO(n) = (O(n))_0$. The fact that the set of orthogonal matrices with the determinant -1 is homeomorphic to $SO(n)$ completes the proof of the lemma.

Sometimes, it is convenient to represent $O(n)$ as a subset in $\mathbf{R}^{n^2}$ defined by the system of equations $AA^T = E$, where $\mathbf{R}^{n^2}$ is identified with the linear space of matrices A of order n.

Consider in $\mathbf{R}^{n^2}$ the form $\langle \mathbf{A}, \mathbf{B}\rangle = \operatorname{spur} AB^T$. Clearly, this scalar product is Euclidean in the basis formed by the vectors $\mathbf{e}_{ij}$ (each vector in $\mathbf{R}^{n^2}$ is identified with a matrix of order n, and all elements of the matrix are zero except for the element equal to unity and located at the intersection of the ith row and jth column, $1 \leqslant i, j \leqslant n$). If

$$\mathbf{A} = \sum_{i,j} a_i^j \mathbf{e}_{ij} \quad \mathbf{B} = \sum_{i,j} b_i^j \mathbf{e}_{ij},$$

$$\text{then } \langle \mathbf{A}, \mathbf{B}\rangle = \operatorname{spur} AB^T = \sum_{i,j} a_i^j b_i^j,$$

which, obviously, coincides with the scalar Euclidean product in the orthogonal basis $\{\mathbf{e}_{ij}\}$. Identifying each matrix $A \in \mathbf{R}^{n^2}$ with the vec-

tor $\mathbf{A} = \sum_{ij} a_i^j \mathbf{e}_{ij}$, we can associate to this matrix the Euclidean length of the vector $\mathbf{A}$, where $\| A \|^2 = \operatorname{spur} AA^T$. As was already noted, $O(n)$ is realized as a subset in $\mathbf{R}^{n^2}$ satisfying the equation $AA^T = E$. Hence, for $A \in O(n)$ we have $\| A \| = \sqrt{n}$, i.e. $O(n)$ is located in a standard sphere $S^{n^2-1} \subset \mathbf{R}^{n^2}$ of radius $\sqrt{n}$ (Fig. 4.60).

4. **Unitary group $U(n)$ and special unitary group $SU(n)$.** Let us consider a complex space $\mathbf{C}^n$ referred to the coordinates $z^1, \ldots, z^n$ and provide it with the Hermitian scalar product $\langle \mathbf{a}, \mathbf{b} \rangle = \operatorname{Re} \sum_{i=1}^{n} a^i \overline{b^j}$ associated with the bilinear complex-valued form $\sum_{i=1}^{n} a^i \overline{b^j}$. Let $U(n)$ stand for the group of all linear operators in $\mathbf{C}^n$ that preserve this scalar product, i.e. the group of all complex-valued matrices A of order n such that $\langle \mathbf{a}, \mathbf{b} \rangle \equiv \langle A\mathbf{a}, A\mathbf{b} \rangle$ for

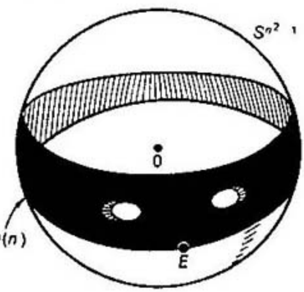

Figure 4.60

any $\mathbf{a}, \mathbf{b} \in \mathbf{C}^n$. This condition is equivalent to the matrix equation $A\overline{A}^T = E$, where the bar denotes a complex conjugate quantity. Thus, if $g \in U(n)$, then $\det g = e^{i\varphi}$. Define $SU(n)$ as a subgroup in $U(n)$ such that $\det g = 1$.

Lemma 11. *The groups $U(n)$ and $SU(n)$ are pathwise connected, i.e. coincide with their unit components.*

Proof. It is known from algebra that for any $g_0 \in SU(n)$ there exists a unitary transformation $g \in U(n)$ such that $a = gg_0g^{-1}$ is a diagonal matrix of the form

$$\begin{pmatrix} e^{i\varphi_1} & & 0 \\ & e^{i\varphi_2} & \\ & & \ddots & \\ 0 & & & e^{i\varphi_n} \end{pmatrix}, \quad \varphi_1 + \ldots + \varphi_n = 2l\pi, \quad l \in \mathbf{Z}$$

(here $e^{i\varphi_k}$, $1 \leqslant k \leqslant n$, are the eigenvalues of the operator g_0). As regards the continuous path $\gamma(t)$ connecting g_0 and E, it is sufficient to consider the continuous family of matrices $\gamma(t) = g^{-1}a(t)\,g$, where

$$a(t) = \begin{pmatrix} e^{i\varphi_1 t} & & & \\ & \ddots & & 0 \\ & & e^{i\varphi_{n-1}t} & \\ & 0 & & e^{i\left(2l\pi - t\sum\limits_{k=1}^{n-1}\varphi_k\right)} \end{pmatrix}.$$

Apparently, $\gamma(0) = E$ and $\gamma(1) = g_0$. That $U(n)$ is connected can be proved in a similar way, and for $\gamma(t)$ we may take the path $\gamma(t) = g^{-1}a(t)\,g$, where

$$a(t) = \begin{pmatrix} e^{i\varphi_1 t} & & & \\ & \ddots & & 0 \\ 0 & & e^{i\varphi_{n-1}t} & \\ & & & e^{i\varphi_n t} \end{pmatrix}.$$

The lemma is proved.

Problem. Demonstrate that $U(n)$ (as a topological space) is isomorphic to the direct product of $SU(n)$ and S^1.

Problem. Let $\mathfrak{G}$ be a connected Lie group and let H be its discrete normal subgroup. (The subgroup H is discrete if the unit element of $\mathfrak{G}$ has an open neighbourhood U such that it contains only one element of H, the unit element.) Prove that any discrete normal subgroup H lies at the centre of $\mathfrak{G}$, i.e. commutes with the whole group $\mathfrak{G}$.

It is sometimes convenient to represent $U(n)$ as a subset in $\mathbf{R}^{2n^2}$ defined by the system of equations $A\overline{A}^T = E$, where $\mathbf{R}^{2n^2} \cong \mathbf{C}^{n^2}$ is identified with the linear space of all complex-valued matrices A of order n. Consider in $\mathbf{C}^{n^2}$ the scalar product $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{Re\ spur}\, A\overline{A}^T$ and consider in $\mathbf{R}^{2n^2}$ the basis consisting of matrices $\boxed{\;1\;}$, $\boxed{\;i\;}$, where all the elements are zero except for one located in the kth column and jth row, $1 \leqslant k, j \leqslant n$. If $\mathbf{A}, \mathbf{B} \in \mathbf{C}^{n^2}$, then $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{Re\ spur}\, A\overline{B}^T = \sum\limits_{i,j=1}^{n} a_i^j \overline{b}_i^j$, which coincides with the Hermitian scalar product in $\mathbf{C}^{n^2}$ identified with $\mathbf{R}^{2n^2}$. Assuming each matrix A to be a vector, we can associate with it the Euclidean length of the vector $\mathbf{A}$: $\|\mathbf{A}\|^2 = \mathrm{Re\ spur}\, A\overline{A}^T$. Hence, $U(n)$ lies inside the sphere S^{2n^2-1} of radius $\sqrt{n}$.

Now that we have demonstrated pathwise connectedness of $SO(n)$ and $SU(n)$, we can prove the following statement.

Lemma 12. *The group $\mathfrak{G}L(n, \mathbf{R})$ consists of two path components. The group $\mathfrak{G}L(n, \mathbf{C})$ is pathwise connected.*

Proof. The group $\mathfrak{G}L(n, \mathbf{R})$ splits into the union of two subsets, $\mathfrak{G}_0 = \{g: \det g > 0\}$ and $\mathfrak{G}_1 = \{g: \det g < 0\}$. These subsets are disjoint since the determinant of the matrix is a smooth function of its coefficients, i.e. of the coordinates in $\mathbf{R}^{n^2}$. Also, $\mathfrak{G}_1$ is homeomorphic to $\mathfrak{G}_0$, the homeomorphism being realized by the mapping

$g \to ag$, where $a = \begin{pmatrix} 1 & & & \\ & \ddots & & 0 \\ 0 & & 1 & \\ & & & -1 \end{pmatrix}$. It remains to prove that $\mathfrak{G}_0$ is

connected. Since each element $g \in \mathfrak{G}_0$ can be interpreted as a basis in $\mathbf{R}^n$, application of the familiar procedure of basis orthogonalization shows that g can be represented as $g = \alpha \cdot \varphi$, where $\alpha \in SO(n)$ and φ is a top-triangular matrix. The matrix g can now be deformed continuously into the matrix α; to this end, it is sufficient to consider the continuous path $\gamma(t) = \alpha \cdot \varphi(t)$, where

$$\varphi(t) = \begin{pmatrix} \varphi_1 t + (1-t) & & & \\ & \varphi_n t + (1-t) & & \varphi_{ij} t \\ 0 & & \ddots & \\ & & & \varphi_n t + (1-t) \end{pmatrix}.$$

Connectedness of $SO(n)$ proves the lemma for $\mathfrak{G}L(n, \mathbf{R})$. Let us consider $\mathfrak{G}L(n, \mathbf{C})$ and construct a smooth mapping $f: \mathfrak{G}L(n, \mathbf{C}) \to S^1$ setting $f(g) = \det g$. The image of $\mathfrak{G}L(n, \mathbf{C})$ is a circle, and the inverse image of the unit element on S^1 is a subset of matrices in $\mathfrak{G}L(n, \mathbf{C})$ such that $\det(g) = 1$, i.e. $f^{-1}(1) = SL(n, \mathbf{C})$. Since a circle is pathwise connected, it suffices to prove pathwise connectedness of the group $SL(n, \mathbf{C})$. Applying the orthogonalization, which transforms an arbitrary unimodular complex basis into a Hermitian one, we pass, as before, from the connectedness of $SL(n, \mathbf{C})$ to the connectedness of $SU(n)$; the latter group is connected by Lemma 11. Thus, Lemma 12 is proved.

Remark. Connectedness of $\mathfrak{G}L(n, \mathbf{C})$ and disconnectedness of $\mathfrak{G}L(n, \mathbf{R})$ can be explained (not quite rigorously) as follows: the group $\mathfrak{G}L(n, \mathbf{R})$ is obtained from $\mathbf{R}^{n^2}$ by removing the hypersurface $\det(g) = 0$ which "subdivides" $\mathbf{R}^{n^2}$ into two domains, one of them being precisely the connectedness component of the unit element of $\mathfrak{G}L(n, \mathbf{R})$. In the complex case, $\mathfrak{G}L(n, \mathbf{C})$ is obtained from $\mathbf{C}^{n^2} \cong \mathbf{R}^{2n^2}$ by removing the subset $\det g = 0$ of codimension two

(from the real point of view), since the complex equation $\det g = 0$ is equivalent to two real equations, $\mathrm{Re}(\det g) = 0$ and $\mathrm{Im}(\det g) = 0$. It is clear intuitively that a surface of "codimension two" does not subdivide $\mathbf{R}^{2n^2}$ into two parts.

We have defined $U(n)$ as a group of matrices which preserve the real-valued scalar product $\langle \mathbf{a}, \mathbf{b}\rangle = \mathrm{Re} \sum_{k=1}^{n} a^k \overline{b}^k$. However, besides this scalar product, there exists in $\mathbf{C}^n$ an associated bilinear complex-valued form $(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{n} a^k \overline{b}^k$ which, naturally, leads to the group of matrices $U(n)'$ that preserve this form, i.e. satisfy the identity $(B\mathbf{a}, B\mathbf{b}) = (\mathbf{a}, \mathbf{b})$ for any $\mathbf{a}, \mathbf{b} \in \mathbf{C}^n$. A question arises: do $U(n)$ and $U(n)'$ coincide?

Lemma 13. *The group $U(n)$ and the group $U(n)'$ coincide.*

Proof. Since $U(n)'$ preserves $(\mathbf{a}, \mathbf{b})$, it also preserves (separately) the real and imaginary parts of $(\mathbf{a}, \mathbf{b})$, and since $\langle \mathbf{a}, \mathbf{b}\rangle = \mathrm{Re}(\mathbf{a}, \mathbf{b})$, preservation of the form $(\mathbf{a}, \mathbf{b})$ implies preservation of $\langle \mathbf{a}, \mathbf{b}\rangle$. Thus, $U(n)' \subset U(n)$. Conversely, let $A \in U(n)$, i.e. $\langle A\mathbf{a}, A\mathbf{b}\rangle = \langle \mathbf{a}, \mathbf{b}\rangle$ for any $\mathbf{a}, \mathbf{b} \in \mathbf{C}^n$. Since this equality holds true for any $\mathbf{a}, \mathbf{b} \in \mathbf{C}^n$ it is also valid for a pair of vectors $i\mathbf{a}$ and $\mathbf{b}$, i.e. $\mathrm{Re}(i\mathbf{a}, \mathbf{b}) = \mathrm{Re}(A(i\mathbf{a}), A\mathbf{b})$. Since the operator A is complex, $A(i\mathbf{a}) = iA\mathbf{a}$ and hence $\mathrm{Im}(\mathbf{a}, \mathbf{b}) = \mathrm{Im}(A\mathbf{a}, A\mathbf{b})$, that is, the imaginary part of the form $(\mathbf{a}, \mathbf{b})$ is preserved. Thus, A preserves $(\mathbf{a}, \mathbf{b})$, i.e. $U(n) \subset U(n)'$. The lemma is proved.

In what follows we will not distinguish between these two invariance groups.

We now discuss the operation of "making real" which permits the identification of $\mathbf{C}^n$ with $\mathbf{R}^{2n}$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a Hermitian basis in $\mathbf{C}^n$, this means that any vector $\mathbf{z} \in \mathbf{C}^n$ admits a unique decomposition $\mathbf{z} = z^1\mathbf{e}_1 + \ldots + z^n\mathbf{e}_n$, where $\mathbf{z}^k = x^k + iy^k$, x^k, $y^n \in \mathbf{R}$. Consider the set of orthogonal (with respect to $\langle \mathbf{a}, \mathbf{b}\rangle$) vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n, i\mathbf{e}_1, \ldots, i\mathbf{e}_n$. Then for any $\mathbf{z} \in \mathbf{C}^n$ there exists the decomposition $\mathbf{z} = \sum_{k=1}^{n} x^k\mathbf{e}_k + \sum_{k=1}^{n} y^k(i\mathbf{e}_k)$, where x^k, $y^k \in \mathbf{R}$. This enables identification of $\mathbf{C}^n$ with $\mathbf{R}^{2n}$. The mapping $\varphi\colon \mathbf{C}^n \to \mathbf{R}^{2n}$ defined by the formula $\varphi(\mathbf{z}) = (x^1, \ldots, x^n, y^1, \ldots, y^n)$ will be called the operator of "making $\mathbf{C}^n$ real". What does it happen to the Hermitian form $(\mathbf{a}, \mathbf{b})$ under φ? Clearly,

$$(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{n} a^k\overline{b}^k \to \sum_{k=1}^{n} (x^k + iy^k)(c^k - id^k)$$

$$= \sum_{k=1}^{n} (x^k c^k + y^k d^k) + i \sum_{k=1}^{n} (y^k c^k - x^k d^k),$$

i.e.

$$\langle \mathbf{a}, \mathbf{b} \rangle = \operatorname{Re}(\mathbf{a}, \mathbf{b}) \to \sum_{k=1}^{n} (x^k c^k + y^k d^k).$$

In other words, the Hermitian scalar product in $\mathbf{C}^n$ becomes the Euclidean scalar product in $\mathbf{R}^{2n}$.

What do we mean by saying: "define in $\mathbf{R}^{2n}$ a complex structure"? Consider the operation $\varphi\colon \mathbf{C}^n \to \mathbf{R}^{2n}$, then an orthogonal basis $\{\mathbf{e}_k, i\mathbf{e}_k\}$, $1 \leqslant k \leqslant n$, arises in $\mathbf{R}^{2n}$, and $\mathbf{R}^{2n} \approx \mathbf{R}^n \oplus i\mathbf{R}^n = \mathbf{C}^n$. Thus, in $\mathbf{R}^{2n}$ there arises a linear operator A such that $A(\mathbf{x}) = i\mathbf{x}$. Clearly, $A^2 = -E$ and $A(i\mathbf{e}_k) = -\mathbf{e}_k$, i.e. in the orthogonal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_k, i\mathbf{e}_1, \ldots, i\mathbf{e}_k)$ the matrix A is of the form $A = \begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix}$, where E is the identity matrix of order n. This operator has $2n$ eigenvalues $\lambda_1 = \ldots = \lambda_n = +i$, $\lambda_{n+1} = \ldots = \lambda_{2n} = -i$ (verify!) and A is also orthogonal: $A \in SO(2n)$.

Definition. We say that an *orthogonal operator* $A \in SO(2n)$ in $\mathbf{R}^{2n}$, such that $A^2 = -E$, *defines a complex structure* in $\mathbf{R}^{2n}$.

This definition will be justified if we can identify $\mathbf{R}^{2n}$ with $\mathbf{C}^n$ via A. Since the operator A is orthogonal, it can be transformed, by an orthogonal rotation of the basis, into

$$A = \left(\begin{array}{cc|c|cc} \cos\varphi_1 & \sin\varphi_1 & & & \\ -\sin\varphi_1 & \cos\varphi_1 & & 0 & \\ \hline & & \ddots & & \\ \hline & 0 & & \cos\varphi_n & \sin\varphi_n \\ & & & -\sin\varphi_n & \cos\varphi_n \end{array}\right).$$

Since $A^2 = -E$, it follows that

$$\varphi_k = \frac{\pi}{2} + l_k \pi, \quad l_k \in \mathbf{Z}.$$

Under another basis transformation (permutations), we can represent A in the form $A = \begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix}$. Thus, we have obtained an orthogonal basis $\mathbf{e}_1, \ldots, \mathbf{e}_n, \mathbf{t}_1, \ldots, \mathbf{t}_n$ such that $A(\mathbf{e}_k) = \mathbf{t}_k$, $A(\mathbf{t}_k) = -\mathbf{e}_k$. Spanning the space $\mathbf{R}^n\{\mathbf{e}_k\}$ by $\mathbf{e}_1, \ldots, \mathbf{e}_n$, we obtain the decomposition $\mathbf{R}^{2n} = \mathbf{R}^n\{\mathbf{e}_k\} \oplus \mathbf{R}^n\{\mathbf{t}_k\}$ such that $A\colon \mathbf{R}^n\{\mathbf{e}_k\} \to \mathbf{R}^n\{\mathbf{t}_k\}$, $A\colon \mathbf{R}^n\{\mathbf{t}_k\} \to \mathbf{R}^n\{\mathbf{e}_k\}$. Hence, any $\mathbf{a} \in \mathbf{R}^{2n}$ admits the unique writing $\mathbf{a} = \mathbf{x} + A\mathbf{y}$, where $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n\{\mathbf{e}_k\}$. Since $A^2 = -E$,

we obtain $\mathbf{C}^n$ by considering all linear combinations of the vectors $\{\mathbf{e}_k\}$.

How many complex structures can be defined in $\mathbf{R}^{2n}$? In other words, how can we describe all $A \in SO(2n)$ such that $A^2 = -E$? Since $A = -A^{-1}$ and $A^{-1} = A^T$, then $A^T = -A$, i.e. A is a complex structure if and only if this operatior is skew-symmetric.

Let us consider the transformation $\varphi\colon \mathbf{C}^n \to \mathbf{R}^{2n}$ and let $\Pi^2 \subset \mathbf{R}^{2n}$ be an arbitrary two-dimensional real plane. When is Π^2 a complex straight line, i.e. the image of a complex straight line under φ? Apparently Π^2 is a complex straight line if and only if it is an invariant plane for the operator $A\colon \mathbf{R}^{2n} \to \mathbf{R}^{2n}$. The proof is left to the reader.

What does it happen to a unitary group under the transformation φ: $\mathbf{C}^n \to \mathbf{R}^{2n}$? Let $A \in U(n)$; since this operator preserves the Hermitian form, φ transforms A into the operator φA which acts in $\mathbf{R}^{2n}$ and preserves the Euclidean form because the Hermitian form is transformed into the Euclidean one. Since φA preserves the Euclidean form, it is an element of the group $O(2n)$. What is the explicit form of the embedding $\varphi\colon U(n) \to O(2n)$?

Proposition 2. *The monomorphism $\varphi\colon U(n) \to SO(2n)$ that arises under the operation $\varphi\colon \mathbf{C}^n \to \mathbf{R}^{2n}$ can be written as*

$$A = C + iB \xrightarrow[\varphi]{} \begin{pmatrix} C & -B \\ B & C \end{pmatrix} \in SO(2n),$$

where C and B are real. Furthermore, $\varphi U(n) = SO(2n) \cap \varphi GL(n, \mathbf{C})$.

Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis in $\mathbf{C}^n$, then $\varphi\colon \mathbf{C}^n \to \mathbf{R}^{2n}$ transforms this basis into $\mathbf{e}_1, \ldots, \mathbf{e}_n, i\mathbf{e}_1, \ldots, i\mathbf{e}_n$, whence $(C + iB)\mathbf{e}_k = C\mathbf{e}_k + B(i\mathbf{e}_k)$, $(C + iB)(i\mathbf{e}_k) = -B\mathbf{e}_k + C(i\mathbf{e}_k)$, i.e. $\varphi A = \begin{pmatrix} C & -B \\ B & C \end{pmatrix}$. Straightforward calculation shows that $\det(\varphi A) = |\det A|^2$, i.e. $\det(\varphi A) > 0$. We now prove that $\varphi U(n) = SO(2n) \cap \varphi GL(n, \mathbf{C})$. Let $\varphi A \in \varphi U(n)$, then, on the one hand, $\varphi A \in SO(2n)$ (see above), and on the other hand, φA is obtained when the transformation is applied to a complex-valued non-degenerate operator, i.e. $\varphi A \in \varphi GL(n, \mathbf{C})$, whence $\varphi U(n) \subset SO(2n) \cap \varphi GL(n, \mathbf{C})$. Conversely, let $g \in SO(2n)$ and $g \in \varphi GL(n, \mathbf{C})$. This means that g is of the form $\begin{pmatrix} C & -B \\ B & C \end{pmatrix}$, i.e. $g \in \varphi U(n)$. The proposition is proved.

There exists another obvious embedding $U(n) \supset SO(n)$ as a subgroup of real matrices.

If $A \in U(n)$ is the operator of multiplication by i, i.e. $A\mathbf{z} = i\mathbf{z}$, the transformation φ sends this operator into $\begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix}$. We have

demonstrated above that the set of complex structures in $\mathbf{R}^{2n}$ coincides with the set of all orthogonal skew-symmetric matrices. If $A \in SO(2n)$ is a complex structure, there exists a matrix $C \in SO(2n)$ such that $A = CT_0C^{-1}$, where

$$T_0 = \left[\begin{array}{cc|c|cc} 0 & 1 & & & \\ -1 & 0 & & 0 & \\ \hline & & \ddots & & \\ \hline & 0 & & 0 & 1 \\ & & & -1 & 0 \end{array}\right].$$

The converse statement is also valid: any operator of the form CT_0C^{-1} is a complex structure. Thus, the set of all complex structures in $\mathbf{R}^{2n}$ can be identified with the set of matrices of the form CT_0C^{-1}, where $C \in SO(2n)$.

It can be proved (the proof is omitted) that all the matrix groups considered above are Lie groups.

5. **Symplectic group** $Sp(n)$ is defined in terms of the algebra of quaternions $\mathbf{Q}$. Recall the definition of $\mathbf{Q}$. Consider $\mathbf{R}^4$ referred to an orthogonal basis whose vectors are denoted by $1, i, j, k$, so that any $q \in \mathbf{R}^4$ is written as $q = a^0 \cdot 1 + a^1 \cdot i + a^2 \cdot j + a^3 \cdot k$, where $a^0, a^1, a^2, a^3 \in \mathbf{R}$. Multiplication in $\mathbf{R}^4$ is defined on the basis 1, i, j, k and is extended then to all vectors in $\mathbf{R}^4$. The table is of the form

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

Thus, we obtain a four-dimensional algebra over $\mathbf{R}$ which is associative but not commutative. This algebra is called the algebra of quaternions $\mathbf{Q}$. There exists in $\mathbf{Q}$ the operation of conjugation: $q \to \bar{q} = a^0 \cdot 1 - a^1 \cdot i - a^2 \cdot j - a^3 \cdot k$ and the operation of finding the inverse element (for a non-zero element): $q^{-1} = \bar{q}/|q|^2$, where

$|q|^2 = q\cdot\bar{q} = \sum_{i=0}^{3} (a^i)^2$. The algebra Q admits the scalar real-valued product $\langle q_1, q_2\rangle = \mathrm{Re}\,(q_1\cdot\bar{q}_2)$, where $(q_1\cdot q_2)$ denotes the product in Q. The elements $q \in \mathrm{Q}$ are called *quaternions*; the quaternions with $a^0 = 0$ are called imaginary. The coordinate a^0 in the decomposition $q = a^0\cdot 1 + a^1\cdot i + a^2\cdot j + a^3\cdot k$ is denoted by $a^0 = \mathrm{Re}\,(q)$; thus, $q = \mathrm{Re}\,(q) + \mathrm{Im}\,(q)$.

We now consider an n-dimensional quaternion space Q^n with the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$; any vector of this space $\mathbf{a} \in \mathrm{Q}^n$ admits a unique representation in the form $\mathbf{a} = \sum_{k=1}^{n} q^k\mathbf{e}_k$, where $q^k \in \mathrm{Q}$.

Definition. A *symplectic group* $Sp\,(n)$ is the set of all linear quaternion transformations of Q^n preserving the point O and leaving invariant the following scalar poduct in Q^n:

$$\langle \mathbf{a}, \mathbf{b}\rangle = \mathrm{Re}\,(\sum_{k=1}^{n} a^k\bar{b}^k),$$

$$\text{i.e. } Sp\,(n) = \{A\colon \mathrm{Q}^n \to \mathrm{Q}^n,\ \langle \mathbf{a}, \mathbf{b}\rangle = \langle A\mathbf{a}, A\mathbf{b}\rangle,\ \mathbf{a}, \mathbf{b} \in \mathrm{Q}^n\}.$$

The number n is called the *quaternion dimension* of Q^n. The space Q^n can be identified canonically with C^{2n}. Take $n = 1$; then $\mathrm{Q}^1 = \mathrm{Q}$. Let $q = a^0\cdot 1 + a^1\cdot i + a^2\cdot j + a^3\cdot k$; using the multiplication table, we can write q in the form $q = (a^0 + a^1 i) + j\,(a^2 - a^3 i) = z^1 + j\bar{z}^2$ where $z^1 = a^0 + a^1 i$, $z^2 = a^2 + a^3 i$ are complex numbers. Performing this operation along each quaternion coordinate in Q^n, we obtain the identification $\mathrm{Q}^n \approx \mathrm{C}^{2n}$. As in the complex case, there exisis, along with $Sp\,(n)$, the invariance group of the quaternion-valued form

$$(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{n} a^k\bar{b}^k, \quad a^k, b^k \in \mathrm{Q}.$$

The following statement is valid: the invariance group of $(\mathbf{a}, \mathbf{b})$ coincides with the group $Sp\,(n)$. The proof is left to the reader as a useful exercise.

Let us consider the operation of "making complex", i.e. identification of Q^n with C^{2n} (not to be confused with complexification!). How does this operation change a quaternion-valued form $(\mathbf{a}, \mathbf{b})$? Put $n = 1$, then

$$a\cdot\bar{b} \to (p + j\bar{q})\,\overline{(c + j\bar{d})} = (p + j\bar{q})\cdot(\bar{c} - dj) = (p\bar{c} - j\bar{q}\,dj)$$
$$- (p\,dj - j\bar{q}\bar{c}) = (p\bar{c} + q\bar{d}) + (-pd + qc)\,j,$$

where $a \to p + j\bar{q}$, $b \to c + j\bar{d}$ (see above). We have used the relations $jq = \bar{q}j$, $j^2 = -1$, and $\overline{a \cdot b} = \bar{b} \cdot \bar{a}$ (verify!). For an arbitrary n the form $(\mathbf{a}, \mathbf{b})$ becomes

$$\sum_{k=1}^{n} (p^k \bar{c}^k + q^k \bar{d}^k) + \sum_{k=1}^{n} (q^k c^k - p^k d^k)\, j,$$

where

$$\mathbf{a} = (a^1, \ldots, a^n) \to (p^1 + j\bar{q}^1, \ldots, p^n + j\bar{q}^n),$$
$$\mathbf{b} = (b^1, \ldots, b^n) \to (c^1 + j\bar{d}^1, \ldots, c^n + j\bar{d}^n).$$

Apparently, the form $(\mathbf{a}, \mathbf{b})_{\mathrm{Herm}} = \sum_{k=1}^{n} (p^k \bar{c}^k + q^k \bar{d}^k)$ coincides with the Hermitian form in $\mathbf{C}^{2n}$, and the form $(\mathbf{a}, \mathbf{b})_{\mathrm{sks}} = \sum_{k=1}^{n} (q^k c^k - p^k d^k)$ is skew-symmetric. i.e. $(\mathbf{a}, \mathbf{b})_{\mathrm{sks}} = -(\mathbf{b}, \mathbf{a})_{\mathrm{sks}}$. If the operator $A: \mathbf{Q}^n \to \mathbf{Q}^n$ preserves $(\mathbf{a}, \mathbf{b})_{\mathrm{quat}}$, it will preserve, after the identification of $\mathbf{Q}^n$ with $\mathbf{C}^{2n}$, two forms: $(\mathbf{a}, \mathbf{b})_{\mathrm{Herm}}$ and $(\mathbf{a}, \mathbf{b})_{\mathrm{sks}}$. That the operator preserves $(\mathbf{a}, \mathbf{b})_{\mathrm{Herm}}$ means that it becomes a unitary operator. We have proved that $Sp\,(n)$ can be embedded in $U\,(2n)$ as a subgroup of those elements that preserve in $\mathbf{C}^{2n}$ the skew-symmetric form $(\mathbf{a}, \mathbf{b})_{\mathrm{sks}} = \sum_{k=1}^{n} (q^k c^k - p^k d^k)$. Here $\mathbf{C}^{2n}$ is assumed to be represented in the form $\mathbf{C}^n \oplus \mathbf{C}^n$, i.e. the coordinates in $\mathbf{C}^{2n}$ are divided into two groups: $\mathbf{a} = (p^1, \ldots, p^n, q^1, \ldots, q^n)$.

Problem. Find the explicit formula for the embedding $Sp\,(n) \to U\,(2n)$.

Proposition 3. *The groups $\mathfrak{G}L\,(n, \mathbf{R})$, $\mathfrak{G}L\,(n, \mathbf{C})$, $SL\,(n, \mathbf{R})$, and $SL\,(n, \mathbf{C})$ are non-compact, and the groups $U\,(n)$, $SU\,(n)$, $O\,(n)$, $SO\,(n)$, and $Sp\,(n)$ are compact (as topological spaces).*

Proof. That $\mathfrak{G}L\,(n, \mathbf{R})$ and $\mathfrak{G}L\,(n, \mathbf{C})$ are non-compact follows from the fact that these groups are unbounded Euclidean domains. Compactness of $U\,(n)$, $SU\,(n)$, $O\,(n)$, and $SO\,(n)$ is a direct consequence of the fact that these groups have been realized as bounded subsets in the sphere. Compactness of $Sp\,(n)$ is a consequence of the fact that this group is realized as a closed subset in $U\,(2n)$. The proposition is proved.

6. The list of the Lie groups considered above is far from being complete. There also exist both compact and non-compact Lie groups. For example, among non-compact Lie groups of great importance are the groups $O\,(n, k)$ and $SO\,(n, k)$. $O\,(n, k)$ is the form invariance group (in $\mathbf{R}^n$): $(\mathbf{a}, \mathbf{b}) = -\sum_{i=1}^{k} a^i b^i + \sum_{j=k+1}^{n} a^j b^j$. $SO\,(n, k)$ is the subgroup of unimodular matrices in $O\,(n, k)$.

Let us consider some Lie groups of low dimension. We have seen that $SO(2)$ is homeomorphic to a circle; $SO(3)$ is homeomorphic to $\mathbf{RP}^3$. Clearly, $U(1) \cong S^1$. Let us study $Sp(1)$ and $SU(2)$.

Proposition 4. *The groups $SU(2)$ and $Sp(1)$ are isomorphic (in the algebraic sense) and are both homeomorphic to a sphere S^3. The group $SO(3)$ is a factor group of $SU(2)$ with respect to the subgroup $\mathbf{Z}_2$ consisting of the elements $(E, -E)$.*

Proof. The group $Sp(1)$ acts in $\mathbf{Q}^1 = \mathbf{Q}$ as multiplication by a "scalar", a quaternion, i.e. each $A: \mathbf{Q} \to \mathbf{Q}$, $A \in Sp(1)$ is of the form $Aq = q \cdot \bar{a}$, where $q \in \mathbf{Q}$ and $a \in \mathbf{Q}$ is a fixed quaternion. Since $(q_1, q_2)_{\text{quat}} = q_1 \cdot \bar{q}_2$ is preserved under A, we have $(q_1, q_2)_{\text{quat}} = (Aq_1, Aq_2)_{\text{quat}}$, i.e. $q_1\bar{q}_2 = q_1\bar{a}a\bar{q}_2 = |a|^2 q_1 \cdot \bar{q}_2$, i.e. $|a|^2 = 1$, $|a| = 1$. Since $|a| \cdot |b| = |a \cdot b|$ (verify!), $Sp(1)$ is isomorphic (in the algebraic sense) to the group of all quaternions $a \in \mathbf{Q}$ such that $|a| = 1$. Such quaternions form a sphere S^3 because $|a|^2 = (a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2$. The unity element of the group S^3 is $1 = (1, 0, 0, 0)$. We now demonstrate that $Sp(1)$ is isomorphic to $SU(2)$. Consider the embedding $Sp(n) \to U(2n)$ (see above). Since $n = 1$, this embedding is of the form $\begin{pmatrix} p & q \\ -\bar{q} & \bar{p} \end{pmatrix}$, i.e. the set of such matrices form $SP(1)$ in $U(2)$, and furthermore $p\bar{p} + q\bar{q} = 1$. Indeed, the mapping $A: q \to q \cdot \bar{a}$ results in $q \to qa = (\alpha + j\beta) \times (p + j\bar{q}) = (\alpha p - \beta\bar{q}) + (\alpha q + \beta\bar{p})j$, i.e. the operator matrix in the basis $\mathbf{e}_1 = (\alpha = 1, \beta = 0)$ and $\mathbf{e}_2 = (\alpha = 0, \beta = 1)$ is of the form $\begin{pmatrix} p & -\bar{q} \\ q & \bar{p} \end{pmatrix}$, which is what was required (after the substitution $\bar{q} \to -\bar{q}$). Since each $A \in Sp(1)$ is an isometry of $\mathbf{R}^4$, it preserves the four-dimensional volume, i.e. $\det \begin{pmatrix} p & -\bar{q} \\ q & \bar{p} \end{pmatrix} = 1$ and therefore $|p|^2 + |q|^2 = 1$, which is what we need.

Thus $Sp(1)$ is isomorphic to the unimodular subgroup in $U(2)$, and the dimension of this subgroup is 3. Since the dimension of $SU(2)$ in $U(2)$ is also 3, $Sp(1)$ and $SU(2)$ are isomorphic. We have also proved that the elements of $SU(2)$ can be written as $\begin{pmatrix} p & q \\ -\bar{q} & \bar{p} \end{pmatrix}$, $|p|^2 + |q|^2 = 1$.

We now prove that there exists a homomorphism $f: SU(2) \to SO(3)$ such that it is an epimorphism and has a kernel isomorphic to $\mathbf{Z}_2$. Realize $SU(2)$ as a group of unit quaternions in $\mathbf{Q}$ and put $f(a) = aq\bar{a}$, where $a \in S^3$, $|a| = 1$, $q \in \mathbf{Q}$, and $\operatorname{Re} q = 0$, i.e. q are imaginary quaternions that form $\mathbf{R}^3 \subset \mathbf{R}^4$ orthogonal to $1 \in \mathbf{Q}$. Then f is a homomorphism because $f(a_1 \cdot a_2) = a_1a_2q\bar{a}_2\bar{a}_1 = f(a_1) \cdot f(a_2)$.

The mapping $f(a)$ transforms $\mathbf{R}^3$ (imaginary quaternions) into itself and is an isometry, since $\mathrm{Re}\,(aq_1\bar{a}a\bar{q}_2\bar{a}) = \mathrm{Re}\,(aq_1\bar{q}_2\bar{a}) = \mathrm{Re}\,(q_1\bar{q}_2)$. Thus, $f(a) \in SO(3)$ (the image $f(S^3)$ lies in $SO(3)$ since S^3 is connected). Find the kernel of f. If $aq\bar{a} \equiv q$ for any $q \in \mathbf{R}^3$ ($\mathrm{Re}\, q = 0$), then $aq = qa$, $|a| = 1$, i.e. $a = \pm 1$, since commutation of a with all imaginary quaternions means that the imaginary part of a is zero. Since $\ker f = \mathbf{Z}_2 = (E, -E)$, we have $\dim (S^3/\mathbf{Z}_2) = 3$, and $\dim SO(3) = 3$ implies that f is an epimorphism. The proposition is proved.

That $SO(3) \cong S^3/\mathbf{Z}_2$, corresponds to the representation of $\mathbf{RP}^3$ as the quotient $S^3/\mathbf{Z}_2$, where $\mathbf{Z}_2$ acts on S^3 as multiplication of a vector by -1.

Problems

1. Prove that $O(2)$ coincides with the isometry group of a circle (with the metric $ds^2 = d\varphi^2$).
2. Prove that the group $SL(2, \mathbf{R})/\mathbf{Z}_2$ is pathwise connected.
3. Prove that the group $U(n)$, as a topological space, is homeomorphic to the direct product of $SU(n)$ and a circle S^1.
4. Let $\mathfrak{G}$ be a connected Lie group and let H be a discrete invariant subgroup. Prove that H lies at the centre of $\mathfrak{G}$.
5. Find an explicit form of the embedding $Sp(n) \subset U(2n)$, by analogy with the embedding $U(n) \subset SO(2n)$.

4.4. DYNAMICAL SYSTEMS

We first recall the simplest properties of the gradient of a smooth function. We are already familiar with the vector field $\mathbf{v}(x) = \overrightarrow{\mathrm{grad}}\, f(x)$, where $f(x)$ is a smooth function on $\mathbf{R}^n$. In curvilinear coordinates, however, the gradient is not a vector field, for it is not transformed as a vector, and therefore the gradient can be interpreted as a field if only there exists a Riemannian metric.

The gradient of a function, $\overrightarrow{\mathrm{grad}}\, f$, is related to the derivative of f in the direction of a vector $\mathbf{a} \in \mathbf{R}^n$ by $\frac{df}{d\mathbf{a}} = \langle \mathbf{a}, \overrightarrow{\mathrm{grad}}\, f \rangle$, where $\frac{df}{d\mathbf{a}}$ is the directional derivative, $\langle , \rangle$ is the Euclidean product. Let f be valid on $\mathbf{R}^n$, then this function defines a level hypersurface (we shall speak about a level surface for $n = 3$ and about a level line for $n = 2$). A level hypersurface is described by $n - 1$ parameters (n parameters in $\mathbf{R}^n$ are related by one condition: namely, the equation $f(x) = c = \mathrm{const}$), that is why the dimension of the hypersur-

face $\{f = c\}$ is equal to $n-1$, provided $\{f - c\}$ is a smooth manifold.

Definition. The point $x_0 \in \{f = c\}$ is called a *critical point of the function* f if $\overrightarrow{\text{grad}}\, f(x_0) = 0$; otherwise, x_0 is called a *non-singular* (*non-critical*) point.

Example. Let the function $z = x^2 - y^2$ be defined on $\mathbf{R}^2$; the level line $z = 0$ is of the form $x = \pm y$ and consists of two straight lines. The point 0 is the only critical point of this function.

Let $x_0 \in \{f = c\}$ be a non-singular point. The vector **a** applied at the point x_0 is called *tangent* to the hypersurface $\{f = c\}$ if there exists a smooth curve $\gamma(t)$ entirely belonging to $\{f = c\}$ and such that $\gamma(O) = x_0$ and $\dot{\gamma}(0) = \mathbf{a}$.

Lemma 1. *Let $f(x)$ be a smooth function on $\mathbf{R}^n$ and let $x_0 \in \{f = c\}$ be a non-singular point. Then the vector $\overrightarrow{\text{grad}}\, f(x_0)$ is orthogonal to the hypersurface $\{f = c\}$ at the point x_0, i.e. $\overrightarrow{\text{grad}}\, f(x_0)$ is orthogonal to any vector* **a** *tangent to $\{f = c\}$ at the point x_0.*

Proof. Since $\langle \mathbf{a}, \overrightarrow{\text{grad}}\, f\rangle = \frac{df}{d\mathbf{a}}$, it is sufficient to calculate the derivative of f with respect to **a**, i.e. $\frac{d}{dt} f(\gamma(t))\,|_{t=0}$. Since $f(\gamma(t)) = c = \text{const}$ and $\gamma(t) \subset \{f = c\}$, we have $\frac{df}{d\mathbf{a}} = 0$. The lemma is proved.

If $x_0 \in \{f = c\}$ is a critical point, then $\overrightarrow{\text{grad}}\, f(x_0) = 0$, and we may assume, formally of course, that $\overrightarrow{\text{grad}}\, f(x)$ is orthogonal to $\{f = c\}$ at any point of this hypersurface.

Because the direction and magnitude of $\overrightarrow{\text{grad}}\, f(x)$ show the direction and rate of the increase of f, the function f always grows at a maximal rate along the normal to a level hypersurface.

We recall that a smooth vector field $\mathbf{v}(P)$ is said to be defined on a smooth manifold M^n, if at each point $P \in M^n$ there exists a vector $\mathbf{v}(P) \in T_P M^n$ smoothly dependent on P.

In a local coordinate system $x^1, \ldots, x^n$ a vector field $\mathbf{v}(P)$ can be given by the set of smooth functions $v^i(x^1, \ldots, x^n)$, $1 \leqslant i \leqslant n$. Continuous fields are defined in a similar way.

Definition. The point $P_0 \in M^n$ is called *singular* for a vector field $\mathbf{v}(P)$ if $\mathbf{v}(P_0) = 0$. The singular point $P_0 \in M^n$ of the field $\mathbf{v}(P)$ is called *isolated* if it has an open neighbourhood U in which the field $\mathbf{v}(P)$ does not have singularities other than the point P_0.

Of great importance in physics, however, are discontinuous fields, which are smooth everywhere except at several isolated discontinuity points on M^n. The examples are given below.

Let $\mathbf{v}(P)$ be a smooth field on M^n. We recall that the trajectory $\gamma(t)$ is called an *integral curve* of the field $\mathbf{v}(P)$ if $\dot{\gamma}(t) = \mathbf{v}(\gamma(t))$, i.e. if the tangent velocity vectors to $\gamma(t)$ coincide with the field vectors $\mathbf{v}$.

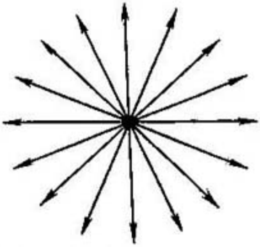

Figure 4.61

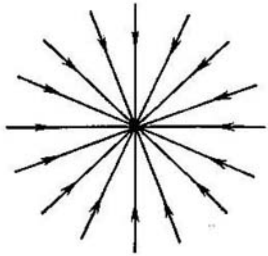

Figure 4.62

Consider several **examples**. Let $f(P)$ be a function on $\mathbf{R}^2$ and $\mathbf{v}(P) = \overrightarrow{\text{grad}}\, f(P)$.

(a) $f(P) = x^2 + y^2$, $\overrightarrow{\text{grad}}\, f = (2x, 2y)$. The integral curves are rays emerging from the point O (Fig. 4.61).

(b) $f(P) = -x^2 - y^2$, $\overrightarrow{\text{grad}}\, f = (-2x, -2y)$. The integral curves are rays converging to the point O (Fig. 4.62).

(c) $f(P) = -x^2 + y^2$, $\overrightarrow{\text{grad}}\, f = (-2x, 2y)$. The integral curves are hyperbolas (Fig. 4.63).

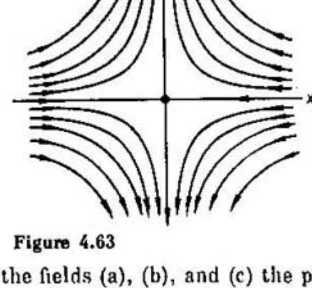

Figure 4.63

For all the fields (a), (b), and (c) the point O is singular. The function f has at O a minimum (case (a)), a maximum (case (b)), and a saddle point (case (c)).

A vector field $\mathbf{v}$ on M^n is frequently meant as the flow of liquid through the manifold and it is assumed that each particle of liquid has a velocity represented by a vector. The singular points of the liquid flow are thus singular points of the field; for example, the singular point in Fig. 4.61 is a source (the liquid "flows out" from the point O), and the singular point in Fig. 4.62 is a sink. The integral curves of the field $\mathbf{v}$ are sometimes called streamlines of a liquid whose flow is described by this velocity field. Not every real liquid flow generates a field in the sense we mean. The fact is that the coordinates of our fields do not depend on time; in other words, liquid flows corresponding to such fields are steady. Time-dependent flows are unsteady. They can be modelled by a family of fields $\mathbf{v}(P, t)$ which smoothly depend on the parameter t.

How can we find explicitly the integral curves of a field? The relevant apparatus has been developed in the theory of ordinary differential equations. Each field $\mathbf{v} = (v^1(P), \ldots, v^n(P))$ can be identified with a system of ordinary differential equations $\frac{dx^k}{dt} = v^k(x^1, \ldots, x^h)$, $1 \leqslant k \leqslant n$. We recall that a system is said to be *autonomous* if its right-hand side does not contain the parameter t explicitly. Steady liquid flows are described by autonomous systems.

Let us consider a particular case where the behaviour of an integral curve can be described in rather simple terms.

Definition. Let, on M^n, be given a system of differential equations describing a field $\mathbf{v}(P)$. A smooth function $f(P)$ on M^n, which is constant along all integral curves of the flow, is called an *integral of the system.*

Let f be an integral of the system. The hypersurfaces $\{f = c\}$ foliate M^n as c varies. Consider a fixed hypersurface $f(P) = c_0$. It follows from the definition just given that if an integral curve has a common point with $\{f = c_0\}$, the trajectory lies entirely in $\{f = c_0\}$. This implies that the field $\mathbf{v}(P)$ is tangent to the surface $\{f = f(P)\}$ at each point P (see Fig. 4.64), that is, each surface $\{f = c\}$ is foliated into integral curves of the flow $\mathbf{v}$. Thus, the order of the initial system of equations can be reduced by one if the field $\mathbf{v}$ is restricted to the surface $\{f = c\}$. We recall that in a neighbourhood of any point, which is non-singular for grad f, the surface $\{f = c\}$ is a smooth manifold of dimension $n - 1$. If we have two functionally independent integrals, f and g (i.e. almost in all points of M^n grad f and grad g are linearly independent), the order of the system is reduced by two units (see Fig. 4.65), etc. If we have a set of $n - 1$ functionally independent integrals, the flow can be integrated completely, i.e. all integral curves can be described as $\gamma(t) = \{f_1 = c_1\} \cap \ldots \cap \{f_{n-1} = c_{n-1}\}$, where $\gamma(0) = P$, $f_k(P) = c_k$, $1 \leqslant k \leqslant n - 1$.

Definition. Let fields $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be given on M^n. These fields are

called *linearly independent* if the vectors $\mathbf{v}_1(P), \ldots, \mathbf{v}_k(P)$ are linearly independent at each point $P \in M^n$.

If some field $\mathbf{v}_\alpha$ vanishes at a point, the entire system $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is no longer independent.

Proposition 1. *Let $M^n = \mathfrak{G}$ be a Lie group. Then there always exist on $\mathfrak{G}$ n linearly independent smooth vector fields $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*

Proof. Suppose $\mathfrak{G}$ is a matrix group and consider on $\mathfrak{G}$ the left translation L_a: $g \to ag$. Clearly, L_a is a diffeomorphism of $\mathfrak{G}$.

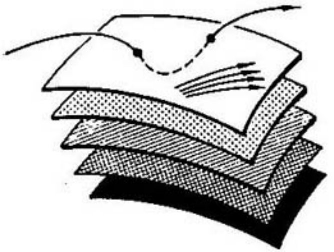

Figure 4.64. The integral curve shown cannot exist.

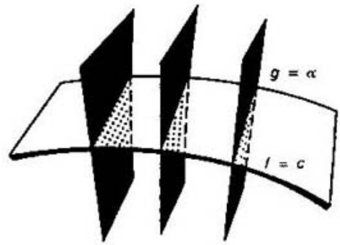

Figure 4.65

Consider the unit element $e \in \mathfrak{G}$. Define n linearly independent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $T_e\mathfrak{G}$ and consider the differential dL_a: $T_e\mathfrak{G} \to T_a\mathfrak{G}$. Put $\mathbf{v}_k(a) = dL_a(\mathbf{e}_k)$. Since L_a is a diffeomorphism, dL_a is non-degenerate, i.e. the vectors $\mathbf{v}_k(a)$, $1 \leqslant k \leqslant n$, are linearly independent.

Below we shall give examples of such M^n on which any smooth field does have singular points (zeros). An interesting example of three independent fields can be constructed on S^3. Since S^3 is homeomorphic to $SU(2)$, there exist three independent fields on S^3, according to Proposition 1. Let us find their explicit form. Let $S^3 = \{q \in \mathbf{Q}, |q| = 1\}$. Put $\mathbf{v}_1(q) = iq$, $\mathbf{v}_2(q) = jq$, and $\mathbf{v}_3(q) = kq$. These fields are tangent to S^3 at the point q. Indeed, calculating the scalar product, say $\langle iq, q\rangle$, we obtain $\langle iq, q\rangle = \mathrm{Re}\,(iq\cdot\bar{q}) = \mathrm{Re}\,(i\,|q|^2) = |q|^2\cdot\mathrm{Re}\,i = 0$.

Problem. Find the integral curves of the above fields on a sphere S^3 in explicit form.

Among smooth fields on M^n of special interest is the class of the so-called gradient or potential fields. Discuss the concept of gradient on M^n in greater detail. Let M^n be provided with a Riemannian metric and let f be a smooth function on M^n. Then, grad $f = \left\{\frac{\partial f}{\partial x^i}\right\}$ is an element of the space dual to $T_P M^n$. To obtain a vector, we consider

(in a local coordinate system) the field $\mathbf{v}(P)$: $v^k(x) = g^{kp}(x)\frac{\partial f(x)}{\partial x^p}$. Calculation shows that the set $v^1, \ldots, v^n$ is transformed exactly in the same way as the coordinates of a vector; this set will be denoted by $\overrightarrow{\text{grad}}\, f$.

Definition. The field of the form $\mathbf{v}(P) = \overrightarrow{\text{grad}}\, f(P)$, where f is a smooth function on M^n, is called a *potential field* on M^n.

Lemma 2. *A potential field does not admit closed integral curves without singular points.*

Proof. Suppose there exists a closed integral curve (such trajectories are sometimes called periodic solutions of a system). Then $\dot{x}(t) = \mathbf{v}(x)$, where $x(t)$ is a solution. If f is the potential, then

$$\frac{df}{d\dot{x}} = \langle \dot{x}, \mathbf{v}(x)\rangle = g_{ij}\dot{x}^i v^j(x) = g_{ij}v^i(x)\, v^j(x) = |\overrightarrow{\text{grad}}\, f|^2 > 0.$$

This inequality shows that f is a strictly monotonic function which grows in the direction of increasing t. But since the trajectory $\gamma(t)$ is closed, the point $\gamma(t)$ will return to the initial position, which contradicts the continuity of f. The lemma is proved.

For example, the field shown in Fig. 4.66 is not potential (the liquid rotates round the origin).

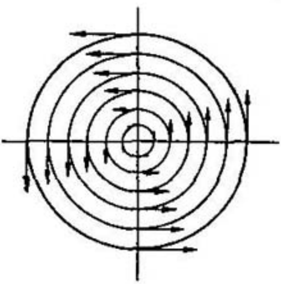

Figure 4.66

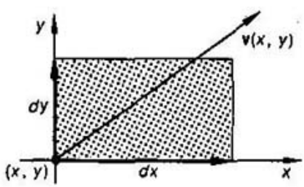

Figure 4.67

Let us consider flows on a two-dimensional manifold. For simplicity, we shall study flows locally; in this case we may consider a flow on a Euclidean plane. Let us treat the flow as a motion of liquid of constant density (equal to 1). We shall study how the mass of liquid at each point changes with time. As is known from the theory of ordinary differential equations, each flow is invariantly associated with a one-parameter group of diffeomorphisms, i.e. shifts along the integral curves of the field. The singular points of a flow are

fixed points of the action of this group. Let the field $\mathbf{v}$ have the coordinates $(P(x, y), Q(x, y))$. Let us consider the change of the mass in an infinitesimal rectangle with sides dx and dy (see Fig. 4.67) (here (x, y) are Cartesian coordinates). If dm is the mass of liquid in the rectangle (dx, dy), then $dm = dx \cdot dy$. Let $\dot{\gamma}(t) = \mathbf{v}(\gamma(t))$ be a system corresponding to $\mathbf{v}$. If $\gamma(t)$ is a solution, t is defined along $\gamma(t)$ to within a shift; we may assume that $\gamma(0) = P$, where P is a point. Set $f_{t_0}(P) = \gamma(t_0)$, where $\gamma(0) = P$. The mapping f_{t_0} is a diffeomorphism. Let us consider $P = (x, y)$ and apply to the rectangle (dx, dy) the inverse mapping $f_{(-\Delta t)}$, where $\gamma(t) = P$. The image of (dx, dy) under $f_{(-\Delta t)}$ is an infinitesimal parallelogram Γ. Applying $f_{\Delta t}$, we transform Γ into (dx, dy) (see Fig. 4.68). The point $P = \gamma(t)$ corre-

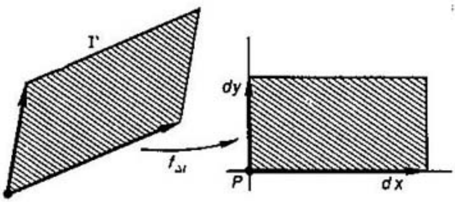

Figure 4.68

sponds to the value t, and the inverse image of P under $f_{\Delta t}$ corresponds to $t - \Delta t$. As t varies, the mass $dm(t) = dx(t) \cdot dy(t)$ also varies.

Lemma 3. *The change of mass in an infinitesimal rectangle* (dx, dy) *can be expressed as*

$$\frac{d}{dt}(dm(t)) = \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\, dy.$$

Proof. We shall assume that $f_{\Delta t}$ defines a new coordinate system in (dx, dy) (transformation by a small diffeomorphism). In the initial coordinate system, $dm(t) = dx(t) \cdot dy(t)$, and in the new coordinate system we have

$$\begin{aligned}
&dx(t+\Delta t) \cdot dy(t+\Delta t) \\
&= d(x(t) + \Delta t \cdot x_t') \cdot d(y(t) + \Delta t \cdot y_t') \\
&= (dx(t) + \Delta t d(x_t')) \cdot (dy(t) + \Delta t d(y_t')) \\
&= (dx(t) + \Delta t dP(x, y)) \cdot (dy(t) + \Delta t dQ(x, y)) \\
&= \left(dx(t) + \Delta t \left(\frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy\right)\right) \\
&\times \left(dy(t) + \Delta t \left(\frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy\right)\right).
\end{aligned}$$

Since in the first factor $dP(x, y)$ is calculated with respect to the increment along the x-axis, the increment Δy in this factor vanishes and only the term $dx(t) + \Delta t \frac{\partial P}{\partial x} dx$ remains. Similarly, in the second factor the increment along the y-axis, i.e. Δx, vanishes, and only the term $dy(t) + \Delta t \frac{\partial Q}{\partial y} dy$ remains. Thus,

$$\left(dx(t) + \Delta t \frac{\partial P}{\partial x} dx\right)\left(dy(t) + \Delta t \frac{\partial Q}{\partial y} dy\right)$$
$$= dx(t) \cdot dy(t) + \Delta t \cdot dx(t)\, dy(t) \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)$$

(small quantities of higher orders are neglected). Hence,

$$\Delta\, dm(t) = \Delta t \cdot \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\, dy,$$

i.e.

$$\frac{d}{dt}(dm(t)) = \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\, dy.$$

The lemma is proved.

Difinition. The function $\operatorname{div} \mathbf{v} = \frac{\partial P(x, y)}{\partial x} + \frac{\partial Q(x, y)}{\partial y}$ is called the *divergence* of a flow $\mathbf{v}$. The flow is incompressible if $\operatorname{div}(\mathbf{v}) = 0$.

Lemma 3 implies that the change in mass is zero if and only if the flow is incompressible. Let us consider an analogue of the expression, obtained in Lemma 3, for a finite domain. Let D be a bounded domain with a piece-smooth boundary and let $D(t)$ be a domain obtained from D as a result of the shift by time t along the integral curves of $\mathbf{v}$ via one-parameter group. Let the area of $D(t)$ be equal to $V(t)$ (if the density of liquid is unity, $V(t)$ is the mass contained in $D(t)$).

Proposition 2. *For any bounded domain $D(t)$*

$$\frac{dV(t)}{dt} = \iint\limits_{D(t)} \operatorname{div} \mathbf{v}\, dx\, dy.$$

Proof. It is sufficient to subdivide $D(t)$ into infinitesimal rectangles and apply Lemma 3.

Definition. The flow $\mathbf{v} = (P, Q)$ is called *irrotational* if $P_y = Q_x$.

This condition can be modified. Let us consider an arbitrary smooth contour C on a plane (a circle without self-intersections), $C = \gamma(t)$; at each point of the contour there arises the number $\langle \mathbf{v}(\gamma(t)), \dot{\gamma}(t)\rangle$ (see Fig. 4.69). The rotor of the flow $\mathbf{v}$ along C is the number $\int\limits_C \langle \mathbf{v}(\gamma(t)), \dot{\gamma}(t)\rangle\, dt$.

Lemma 4. *The flow $\mathbf{v} = (P, Q)$ is irrotational if and only if its rotor along any smooth contour vanishes.*

Proof. Let D be a domain bounded by C. It follows that

$$\int\limits_{\gamma(t)} \langle \mathbf{v}(\gamma(t)),\ \dot{\gamma}(t)\rangle\, dt = \int\limits_{\gamma(t)} P(\gamma(t))\, dx(t) + Q(\gamma(t))\, dy(t)$$
$$= \int\limits_0^{2\pi} \left(P\frac{dx}{dt} + Q\frac{dy}{dt}\right) dt$$
$$= \iint\limits_D (P_y - Q_x)\, dx\, dy, \quad \gamma(0) = \gamma(2\pi).$$

Here we have used the Stokes formula (see Chapter 6). The lemma is proved.

Proposition 3. *An irrotational flow* $\mathbf{v}$ *on a plane is potential, i.e. there exists a smooth function* $a(x, y)$ *such that* $\mathbf{v} = \operatorname{grad} a(x, y)$. *The form* $Pdx + Qdy$ *is the exact differential of the function* $a(x, y)$ *defined uniquely to within an additive constant.*

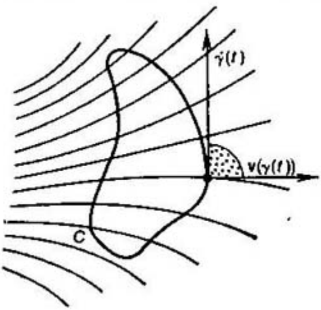

Figure 4.69

Proof. It is sufficient to integrate the system of equations $P = \frac{\partial a}{\partial x}$, $Q = \frac{\partial a}{\partial y}$ under the condition $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$. Integrating the first equation with respect to x, we obtain $a(x, y) = \int\limits_0^x P(x, y)dx +$ $g(y)$. Differentiation with respect to y yields

$$\frac{\partial a(x, y)}{\partial y} = \int\limits_0^x \frac{\partial P(x, y)}{\partial y} dx + \frac{dg(y)}{dy},$$

whence

$$Q(x, y) = \int_0^x \frac{\partial Q(x, y)}{\partial y}\, dx + \frac{dg(y)}{dy},$$

or $Q(x, y) = Q(x, y) - Q(0, y) + \frac{dg(y)}{dy}$, which gives $g'(y) = Q(0, y)$, i.e. $g(y) = \int_0^y Q(0, y)\, dy + C$, where $C = \text{const}$. Thus,

$$a(x, y) = \int_0^x P(x, y)\, dx + \int_0^y Q(0, y)\, dy + C.$$

If we started integration with the equation $Q = \frac{\partial a}{\partial y}$, we would obtain $a(x, y) = \int_0^y Q(x, y)\, dy + \int_0^x P(x, 0)\, dx + C$. The function $a(x, y)$ is the flow potential. Let us describe $a(x, y)$ in geometric terms.

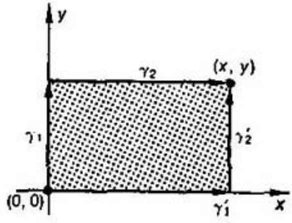

Figure 4.70

Consider two piecewise smooth paths $\gamma = \gamma_1 \cup \gamma_2$ and $\gamma' = \gamma_1' \cup \gamma_2'$ (Fig. 4.70). Clearly,

$$a(x, y) = \int_0^x P(x, y)\, dx + \int_0^y Q(0, y)\, dy = \int_\gamma (P\, dx + Q\, dy),$$

$$a(x, y) = \int_0^y Q(x, y)\, dy + \int_0^x P(x, 0)\, dx = \int_{\gamma'} (P\, dx + Q\, dy),$$

i.e. $a(x, y)$ can be found by integrating the differential form $\omega = Pdx + Qdy$ along either γ or γ' leading from point $(0, 0)$ to point (x, y).

Proposition 4. *Let* $\mathbf{v}$ *be an irrotational flow. Then it is potential and the potential is of the form*

$$a(x, y) = \int_\gamma P\,dx + Q\,dy = \int_\gamma \omega,$$

where γ *is an arbitrary piecewise smooth path from point* $(0, 0)$ *to point* (x, y). *The integral* $\int_\gamma \omega$ *does not depend on the path.*

Proof. We first prove that $\int_\gamma (P\,dx + Q\,dy)$ does not depend on the path (provided the initial and terminal points of the path

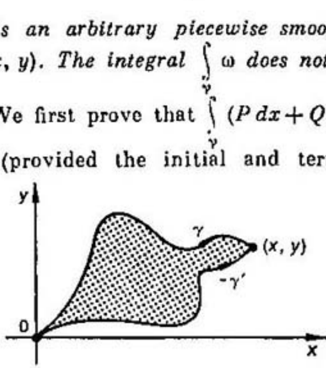

Figure 4.71

are fixed). Let γ' be another path from $(0, 0)$ to (x, y); consider $\alpha = \int_\gamma \omega - \int_{\gamma'} \omega = \int_{\gamma \cup -\gamma'} \omega$, where $-\gamma'$ is the path γ' oriented in the opposite direction (see Fig. 4.71). Then, $\int_{\gamma \cup -\gamma'} (P\,dx + Q\,dy) = \int_C (P\,dx + Q\,dy) = 0$, because $C = \gamma \cup -\gamma'$ is a closed contour and the flow is irrotational (corollary of the Stokes formula). Thus, $\alpha = 0$ and $\int_\gamma \omega = \int_{\gamma'} \omega$. Since we have proved independence of $\int_\gamma \omega$ of the path, it is sufficient to choose one of the paths of Fig. 4.70 to evaluate $\int_\gamma \omega$ numerically. The proposition is proved.

If we choose another initial point of the integration path, $a(x, y)$ changes by an additive constant. Let $\mathbf{v}$ be an irrotational, incompressible flow. Then its coordinates, P and Q, satisfy

$$\frac{\partial P}{\partial x} = -\frac{\partial Q}{\partial y}, \quad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad P = \frac{\partial a}{\partial x}, \quad Q = \frac{\partial a}{\partial y},$$

whence $\frac{\partial^2 a}{\partial x^2}+\frac{\partial^2 a}{\partial y^2}=0$. The linear differential operator of the second order $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$ (written in Cartesian coordinates) is called the *Laplacian*. The function $f(x, y)$, such that $\Delta f=0$, is called *harmonic*.

We have proved that the potential of an irrotational, incompressible flow is a harmonic function. The potential $a(x, y)$ is usually considered together with another potential, $b(x, y)$, called a *conjugate* potential or the potential of a conjugate flow. To define this potential, we consider the system $\frac{\partial b}{\partial x}=-Q$, $\frac{\partial b}{\partial y}=P$. The function $b(x, y)$, which is a solution of this system, is called a conjugate potential. Let us prove the existence of a conjugate potential. Put $\widetilde{P}=-Q$, $\widetilde{Q}=P$. Then, $\frac{\partial b}{\partial x}=\widetilde{P}$, $\frac{\partial b}{\partial y}=\widetilde{Q}$, provided $\frac{\partial \widetilde{Q}}{\partial y}=-\frac{\partial \widetilde{P}}{\partial x}$, $\frac{\partial \widetilde{Q}}{\partial x}=\frac{\partial \widetilde{P}}{\partial y}$. This is the system that was integrated above to find $a(x, y)$. Thus, $b(x, y)$ does exist and plays the role of $a(x, y)$ for the flow $\mathbf{v}=(\widetilde{P}, \widetilde{Q})=(-Q, P)$. The flow $\widetilde{\mathbf{v}}=(\widetilde{P}, \widetilde{Q})$ is called *conjugate* to the flow $\mathbf{v}=(P, Q)$. The converse is also true: the potential $a(x, y)$ is conjugate to $b(x, y)$, i.e. the potential double-conjugate to $a(x, y)$ coincides with $a(x, y)$. The flows $\mathbf{v}$ and $\widetilde{\mathbf{v}}$ are orthogonal: $\langle v, \widetilde{v}\rangle=-PQ+QP=0$.

Let us consider the plane (x, y) as a plane of complex variable $z=x+iy$ and consider also a complex-valued function $f(x, y)=a(x, y)+ib(x, y)$, where a and b are the potential and the conjugate potential of the incompressible flow $\mathbf{v}=(P, Q)$. Instead of $\frac{\partial g}{\partial y}, \frac{\partial g}{\partial x}$ we will simply write g_y, g_x. Since $a_x=P$, $a_y=Q$, $b_x==-Q$, and $b_y=P$, we have $a_x=b_y$, $a_y=-b_x$. The function $f(x, y)=a+ib$ is called complex-analytic, and the equations for a and b are called the Cauchy-Riemann equations (conditions). The functions a and b are called the real and imaginary parts of f, respectively: $a=\operatorname{Re} f$, $b=\operatorname{Im} f$. Let us recall some properties of a complex-analytic function.

Let $z=x+iy$, $\bar{z}=x-iy$, then $x=\frac{1}{2}(z+\bar{z})$, $y=\frac{-i}{2}(z-\bar{z})$. Thus, any function $g(x, y)=u(x, y)+iv(x, y)$, which can be expanded in a convergent series of x and y, can be written as $g(x, y)=\widetilde{g}(z, \bar{z})$. By the rule of differentiation of a composite function,

$$\frac{\partial}{\partial z}=\frac{\partial x}{\partial z}\frac{\partial}{\partial x}+\frac{\partial y}{\partial z}\frac{\partial}{\partial y}=\frac{1}{2}\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)$$

Similarly, $\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$. Among the set of functions $g(z, \bar{z})$ we will choose those that depend only on z. Analytically, this is written as $\frac{\partial}{\partial \bar{z}} g(z, \bar{z}) = 0$. Such functions are called *complex-analytic*, they can only be expanded in a convergent power series of one variable, z. Since $\frac{\partial g}{\partial \bar{z}} = 0$, $g_x + ig_y = 0$, i.e. $u_x + iv_x + i(u_y + iv_y) = 0$, and the condition $\frac{\partial g}{\partial \bar{z}} = 0$ is equivalent to the Cauchy-Riemann equations, $u_x = v_y$, $u_y = -v_x$.

Thus, we have proved

Theorem 1. *Any irrotational, incompressible flow* $\mathbf{v} = (P, Q)$ *can be represented in the form* $\mathbf{v} = \operatorname{grad} a(x, y)$ *and the conjugate flow* $\tilde{\mathbf{v}} = (\tilde{P}, \tilde{Q})$ *is represented in the form* $\tilde{\mathbf{v}} = \operatorname{grad} b(x, y)$, *where* $f(x, y) = a(x, y) + ib(x, y)$ *is a complex-analytic function uniquely defined to within an additive constant. The converse is also true: if* $f(z)$ *is a complex-analytic function, the flows* $\mathbf{v} = \operatorname{grad} \operatorname{Re} f(z)$ *and* $\tilde{\mathbf{v}} = \operatorname{grad} \operatorname{Im} f(z)$ *are irrotational and incompressible, and they are conjugate to each other.*

The integral curves of $\mathbf{v}$ and $\tilde{\mathbf{v}}$ are orthogonal. The function $f = a + ib$ is called the *complex potential* of a flow. We now turn to studying singular points of a vector field on a plane.

Let $\mathbf{v}$ be an irrotational, incompressible flow. How can we find zeros of $\mathbf{v}$ and $\tilde{\mathbf{v}}$? From the Cauchy-Riemann equations we have

$$f'_z(z) = \frac{1}{2}(f_x - if_y) = a_x - ia_y = b_y + ib_x.$$

Lemma 5. *The points at which* $f'_z(z)$ *vanishes coincide with zeros of the flow* $\mathbf{v}$ *(the same points are zeros of* $\tilde{\mathbf{v}}$*).*

The *proof* directly follows from $f'_z(z) = a_x - ia_y = b_y + ib_x$.

Thus, to find zeros of the flows $\mathbf{v}$, $\tilde{\mathbf{v}}$, it suffices to solve the equation $f'(z) = 0$. Until now, we have considered only smooth fields. All the considerations can, nevertheless, be repeated for a flow with isolated discontinuity points which are not, of course, solutions of the equation $f'(z) = 0$.

Here is a question of practical importance: let $\mathbf{v}$ be an incompressible, irrotational flow, i.e. $\mathbf{v} = \operatorname{grad} \operatorname{Re} f(z)$, where $f(z)$ is an analytic function, how can we find integral curves of this field? There is no need to solve explicitly the corresponding system of differential equations.

Proposition 5. *Let* $f = a + ib$ *be a complex analytic function,* $\mathbf{v} = \operatorname{grad} a$, $\tilde{\mathbf{v}} = \operatorname{grad} b$. *Then* b *is an integral for the field* $\mathbf{v}$, *and* a *is*

an integral for $\tilde{\mathbf{v}}$, i.e. the integral curves of $\mathbf{v}$ *are level lines of the function b, and the integral curves of* $\tilde{\mathbf{v}}$ *are level lines of a.*

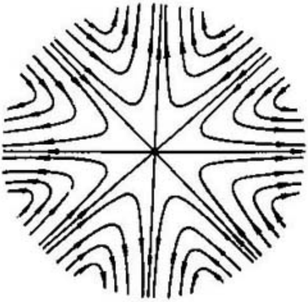

Figure 4.72

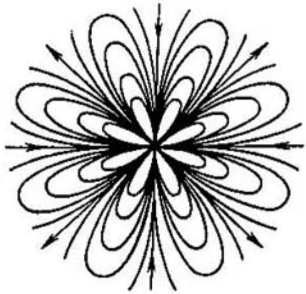

Figure 4.73

Proof. It is sufficient to find $\frac{da}{d\tilde{\mathbf{v}}}$ and $\frac{db}{d\mathbf{v}}$. For example, $\frac{db}{d\mathbf{v}} = \langle \mathbf{v}, \overrightarrow{\text{grad}}\ b\rangle = a_x b_x + a_y b_y = 0$. Similarly, $\frac{da}{d\tilde{\mathbf{v}}} = 0$, that is, a

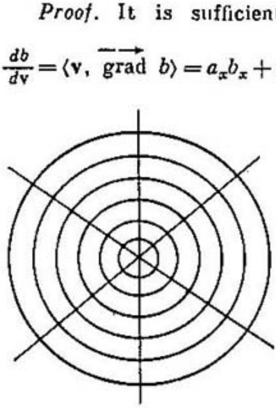

Figure 4.74. The potential $\varphi(x, y)$ is a many-valued function.

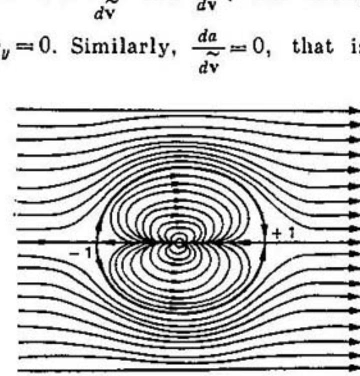

Figure 4.75

and b are constant along the corresponding integral curves. The proposition is proved.

Examples. Let $f(z) = z^k$, $k \geqslant 2$, $f' = kz^{k-1}$. Clearly, $f'(z) = 0$ only at the point 0, $f = r^k (\cos k\varphi + i \sin k\varphi)$, i.e. $a = r^k \cos k\varphi$,

$b = r^k \sin k\varphi$. Figure 4.72 shows the integral curves of grad a (for $k = 4$). The origin is a singular point obtained as a result of the

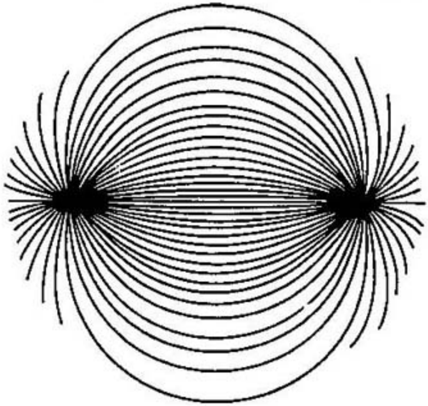

Figure 4.76. Dipole.

confluence of several singularities of lower order (see below). Let $f = z^{-k}$, $k \geqslant 1$, $f = r^{-k} (\cos k\varphi - i \sin k\varphi)$. Figure 4.73 shows the

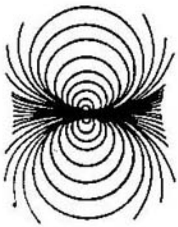

Figure 4.77

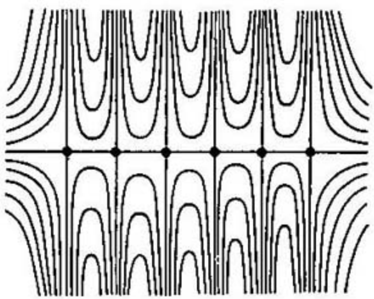

Figure 4.78. $k = 6$.

integral curves of grad a (for $k = 4$). Let $f(z) = \ln z$. The integral curves of $\mathbf{v}$ and $\tilde{\mathbf{v}}$ for this function are shown in Fig. 4.74. This is

a logarithmic singularity. Let us consider the Joukowski transformation, $f(z) = z + 1/z$. The integral curves (for one flow) are presented in Fig. 4.75. **Problem**: Construct trajectories of the conjugate flow.

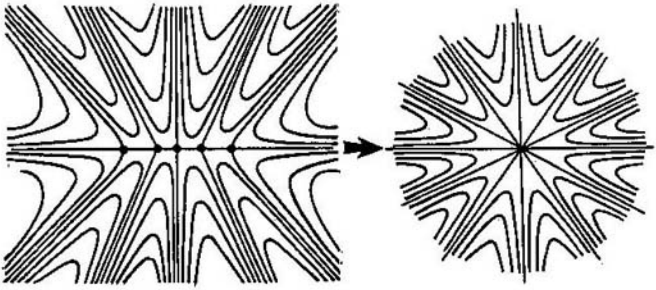

Figure 4.79

Let $f(z) = \frac{1}{\alpha}(\ln(z+\alpha) - \ln(z-\alpha))$, α is real. The pattern of a flow is shown in Fig. 4.76 (construct the pattern of the conjugate flow). This case is an illustration of the confluence of singularities.

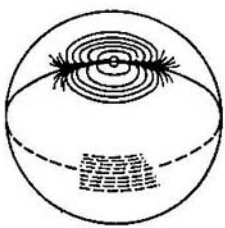

Figure 4.80

For $\alpha \neq 0$ the flow pattern (see Fig. 4.76) coincides with that for a dipole (two charges at singular points). As $\alpha \to 0$, $f(z) \to f'_z(z) = 1/z$ and the dipole field becomes the field of a flow with a first-order pole (see Fig. 4.77).

Let us consider a polynomial $P_k(z) = \prod_{i=1}^{k} (z - \alpha_i)$ with simple real roots. The corresponding flow is shown in Fig. 4.78. Let all the roots α_i vanish (the polynomial is transformed into $\widetilde{P}_k(z) = z^k$ with only one zero of order k). Deformation of the flow is illustrated in Fig. 4.79.

Problem. Construct the deformation of the function z^{-k} to the function $\left(\prod_{i=1}^{k}(z-\alpha_i)\right)^{-1}$ and draw the pattern of flow deformation.

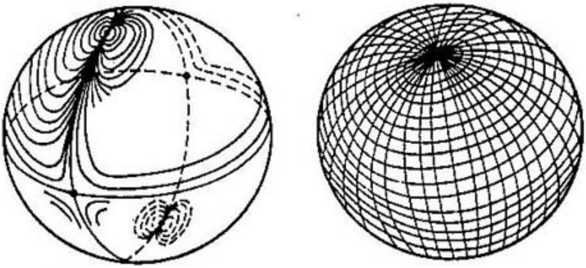

Figure 4.81

Thus far, we have considered flows on a plane. These examples can be transferred to a sphere S^2. The stereographic projection defines a correspondence between the points of S^2 and the points of an

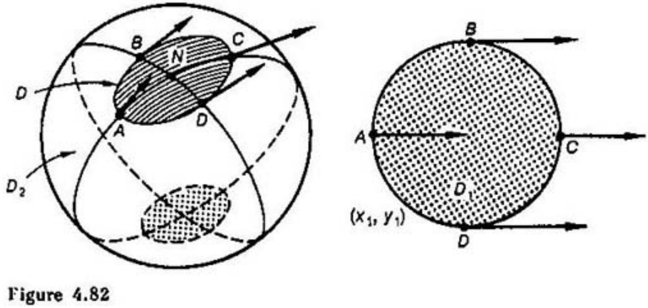

Figure 4.82

extended complex plane (i.e. the complex plane with the adjoined infinitely distant point).

The flow $\mathbf{v}$ = grad Re z on $\mathbf{R}^2$ has only one singular point, infinity, this is a first-order pole (when considered on S^2) (see Fig. 4.80). Figure 4.81 shows the qualitative behaviour of grad Re $(z + 1/z)$, grad Re ln z, and grad Im ln z.

Problem. Construct the pattern of the trajectories for the following flows:

$$(\alpha+i\beta)\cdot\ln\left(\frac{az+b}{cz+d}\right),\quad \frac{1}{z}+\ln z,\quad z^k+\ln z.$$

The attempts at constructing on S^2 a smooth field without singularities have failed; in particular, the flows considered above do have singularities on S^2. This is no accident. Let us explain intuitively why any smooth field on S^2 has a singularity. Suppose a field without singularities is constructed on S^2. Represent S^2 as the union of two disks: $S^2 = D_1 \cup D_2$, where D_1 is the disk of small radius ε with centre at the north pole N, and D_2 is the complement to D_1 (see

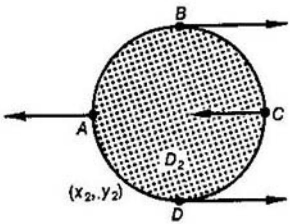

Figure 4.83

Fig. 4.82). Since the field is smooth, we may assume that it is constant in D_1 and its vectors at points A, B, C, D have directions as is shown in the figure. Introduce on D_1 and D_2 Cartesian coordinates (x^1, y^1) and (x^2, y^2), respectively. Then in D_2 the field $\mathbf{v}$ (in these coordinates) has the pattern shown in Fig. 4.83. Such a pattern can be explained as follows. Stretch D_1 along the meridians so that it occupies almost the entire sphere except for D_2 of small radius. Then we obtain the pattern of Fig. 4.83. Since the radius of D_2 is small, it is clear intuitively that this domain should contain a singularity, otherwise, the trajectory pattern would coincide with that in D_1.

4.5. CLASSIFICATION OF TWO-DIMENSIONAL SURFACES

This section deals with the structure of two-dimensional manifolds. We shall start with general definitions, then consider particular examples, and in conclusion give the formal proof of the classification theorem.

4.5.1. MANIFOLDS WITH BOUNDARY

Let us extend the concept of a smooth manifold by including into it subsets in $\mathbf{R}^n$ defined by systems of equations and inequalities. Let $\mathbf{R}^n_+ \subset \mathbf{R}^n$ stand for the half-space given by $x^n \geqslant 0$, $(x^1, \ldots, x^n) \in \mathbf{R}^n$ and let $\mathbf{R}^{n-1}_0$ denote the boundary $x^n = 0$ of $\mathbf{R}^n_+$. If $f\colon \mathbf{R}^n_+ \to \mathbf{R}^1$ is a continuous function, then at all points, including the boundary ones, we shall define differentiability of f. If $\mathbf{x}_0 \in \mathbf{R}^n_+$ is an interior point, i.e. $x^n > 0$, the definition is customary. If $\mathbf{x}_0 \in \mathbf{R}^{n-1}_0$, i.e. $x^n = 0$, we say that f is differentiable at the point $\mathbf{x}_0$ if the following expansion is valid:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} a_i (x^i - x^i_0) + 0(\mathbf{x} - \mathbf{x}_0),$$

where $\lim \frac{0(\mathbf{x}-\mathbf{x}_0)}{|\mathbf{x}-\mathbf{x}_0|} = 0$ for $\mathbf{x} \to \mathbf{x}_0$, $x^n \geqslant 0$. Then $a_i = \frac{\partial f}{\partial x^i}(\mathbf{x}_0)$ for $i = 1, 2, \ldots, n-1$ and $a_n = \lim\limits_{h \to +0} \frac{1}{h}(f(x^1_0, \ldots, x^{n-1}_0, x^n_0 + h) - f(x^1_0, \ldots, x^n_0))$. The latter limit may be called the partial derivative $\frac{\partial f}{\partial x^n}$ of the function f at the point $\mathbf{x}_0$. Smooth functions of class C^r, $r = 1, 2, \ldots, \infty$, are defined in a similar way.

Definition. A metric space M is called a *smooth manifold with boundary* if there exists an atlas $\{U_\alpha\}$ and coordinate homeomorphisms $\varphi_\alpha\colon U_\alpha \to V_\alpha \subset \mathbf{R}^n_+$, where V_α is an open set in $\mathbf{R}^n_+$, such that the functions of coordinate transformation

$$\varphi_\beta \varphi_\alpha^{-1}\colon V_{\alpha\beta} = \varphi_\alpha (U_\alpha \cap U_\beta) \to V_{\beta\alpha} = \varphi_\beta (U_\alpha \cap U_\beta)$$

are smooth functions of class C^r, $r = 1, 2, \ldots, \infty$. The linear coordinates $(x^1, \ldots, x^n)$ in $\mathbf{R}^n_+$ induce local coordinates in the chart U_α: $x^k_\alpha(P) = x^k(\varphi_\alpha P)$. In any local coordinate system we have then $x^n_\alpha(P) \geqslant 0$. A point $P \in M$ is called an *interior point* if $x^n_\alpha(P) > 0$ and it is called a *boundary point*, if $x^n_\alpha(P) = 0$.

If P is an interior point, then $x^n(P) > 0$ in any local coordinate system $(x^1, \ldots, x^n)$, and if P is a boundary point, then in any local coordinate system $(x^1, \ldots, x^n)$ $x^n(P) = 0$. Indeed, suppose that for a point P there exist two coordinate systems $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ such that $x^n(P) > 0$, $y^n(P) = 0$. The coordinates $(x^1, \ldots, x^n)$ define a homeomorphism of some neighbourhood $U \subset M$ into the domain $V \subset \mathbf{R}^n$, and the coordinates $(y^1, \ldots y^n)$ map homeomorphically the same neighbourhood U into the domain $V' \subset \mathbf{R}^n_+$. Consider the transition function as a smooth homeomorphism $\varphi\colon V \to V'$, $y^k = \varphi^k(x^1, \ldots, x^n)$ satisfying the conditions: (a) $y^n = \varphi^n(x^1, \ldots, x^n) \geqslant 0$ and (b) $y^n = \varphi^n(x^1_0, \ldots, x^n_0) = 0$ for a point $\mathbf{x}_0 = (x^1_0, \ldots, x^n_0) \in V$. These conditions mean that at the point $\mathbf{x}_0$ the function φ^n reaches a minimum. Since $\mathbf{x}_0$ is an inte-

rior point for the domain $V \subset \mathbf{R}^n$, the gradient of φ^n vanishes at $\mathbf{x}_0$, i.e.

$$\frac{\partial \varphi^n}{\partial x^i}(x_0^1, \ldots, x_0^n) = 0, \ i = 1, \ \ldots, \ n.$$

Hence, the determinant of the Jacobi matrix $\left\| \frac{\partial y^i}{\partial x^j} \right\|$ at the point $\mathbf{x}_0$ is zero, which contradicts the condition that the transition function from one local coordinate system to another is smooth. Thus, for a manifold with boundary there exists an atlas $\{U_\alpha\}$ with local coordinates $(x_\alpha^1, \ldots, x_\alpha^n)$ such that in any chart we have the strict inequality $x_\alpha^n > 0$ at the interior points and the equality $x_\alpha^n = 0$ at the boundary points. If the chart U_α does not contain boundary points, the condition $x_\alpha^n > 0$ may be rejected without loss of generality. The set ∂M of boundary points is a smooth manifold of dimension $n - 1$. Indeed, the intersection $W_\alpha = \partial M \cap U_\alpha$ can be chosen as a chart and $x_\alpha^1, \ldots, x_\alpha^{n-1} \in U_\alpha$ as coordinates. The coordinate homeomorphisms φ_α (see above) map W_α homeomorphically onto $V_\alpha \cap \mathbf{R}_0^{n-1}$, and the transition functions remain smooth, for they are restrictions of the functions of coordinate transformation to the manifold M.

If $\partial M = \varnothing$, we arrive at the earlier concept of a manifold, henceforth, this manifold will be called closed.

Examples. 1. Consider a ball D^n defined in $\mathbf{R}^n$ by the inequality $\sum_{i=1}^{n} (x^i)^2 \leqslant 1$, the space D^n is a manifold with the boundary S^{n-1}.

2. In general, let $f\colon \mathbf{R}^n \to \mathbf{R}^k$ be a vector function satisfying the conditions of the implicit function theorem, i.e. at all points $P \in M = \{f(P) = 0\}$ the differential df has the maximal rank. Consider the system of equalities and inequalities: $f(x^1, \ldots, x^n) = 0, \ldots, f^{k-1}(x^1, \ldots, x^n) = 0$, and $f^k(x^1, \ldots, x^n) \geqslant 0$ and let M' be the set of solutions of this system. If at each point $P \in M'$ the rank of $\left\| \frac{\partial f^i}{\partial x^s} \right\|$, $1 \leqslant i \leqslant k-1$, $1 \leqslant s \leqslant n$, is equal to $k - 1$, then M' is a manifold with the boundary $\partial M' = M$. Indeed, by the implicit function theorem for the system $f^1 = 0, \ \ldots, \ f^{k-1} = 0$ there exists an atlas at points $P \in M' \setminus M$. Consider a point $P \in M$. Let N' be the set of solutions of the system $f^1 = 0, \ldots, f^{k-1} = 0$ in some neighbourhood of P. Then f^k may be considered as a function on N'. Clearly, the differential of f^k on N' does not vanish. Hence, according to Theorem 1 of Chapter 3 (Sec. 3.5), N', is diffeomorphic to the direct product of the interval $(-\varepsilon, \varepsilon)$ and the manifold N which is the intersection of M and the neighbourhood of the point P, i.e. $N' = N \times (-\varepsilon, \varepsilon)$. Thus, the inequality $f^n \geqslant 0$ defines the

chart $N_1 \times [0, \varepsilon)$ in which the set $(x^1, \ldots, x^{n-k})$ on the manifold N and the function $f^k \geqslant 0$ can be chosen as coordinates.

3. In the preceding example one can see that the Cartesian product of the manifold M with the boundary ∂M and the closed manifold N is the manifold $M \times N$ with the boundary $\partial M \times N$.

4.5.2. ORIENTABLE MANIFOLDS

Definition. A manifold (with boundary) M is called *orientable* if there exists an atlas $\{U_\alpha\}$ such that the Jacobian of the transformation from one local coordinate system to another is positive. We also say that the atlas $\{U_\alpha\}$ for which all Jacobians of the transition functions are positive defines an orientation on M; such an *atlas* is called *oriented*. Two atlases $\{U_\alpha\}$ and $\{U'_\beta\}$ define the same orientation if their union $\{U_\alpha\} \cup \{U'_\beta\}$ is an oriented atlas.

Any atlas on an orientable manifold can be made oriented by the transformation of local coordinates in each chart. Indeed, let $\{U_\alpha, (x^1_\alpha, \ldots, x^n_\alpha)\}$ be an arbitrary atlas consisting of connected charts. Since the manifold M is orientable, there exists on M an oriented atlas $\{V_\beta, (y^1_\beta, \ldots, y^n_\beta)\}$. Let us consider an arbitrary point $P \in U_\alpha$ and let $P \in V_\beta$. If $\frac{\partial(x^1_\alpha, \ldots, x^n_\alpha)}{\partial(y^1_\beta, \ldots, y^n_\beta)}(P) > 0$, we retain in U_α the coordinates $(x^1_\alpha, \ldots, x^n_\alpha)$. If, on the contrary, $\frac{\partial(x^1_\alpha, \ldots, x^n_\alpha)}{\partial(y^1_\beta, \ldots, y^n_\beta)} < 0$, we shall take $(-x^1_\alpha, x^2_\alpha, \ldots, x^n_\alpha)$ as new local coordinates. Thus, we have made the Jacobian of the transition from $(x^1_\alpha, \ldots, x^n_\alpha)$ to $(y^1_\beta, \ldots, y^n_\beta)$ positive at least at one point P of each chart U_α. Since U_α are connected, the Jacobians are positive at any point of U_α.

It is convenient to say that an individual chart defines a local orientation on a manifold M; also, if the Jacobian of the transition from the coordinates in one chart to the coordinates in the other chart is positive, the two local orientations on M are said to be *consistent*. If an orientation is defined on M by an oriented atlas, the local orientation of another connected chart can be consistent or inconsistent with the orientation of M. In the former case we say that the local orientation coincides with the orientation of M, and in the latter case we say that the local orientation is opposite to the orientation of M.

Proposition 1. *On a connected oriented manifold there exist exactly two distinct orientations, any chart defining a local orientation that coincides with one of the orientations of M.*

Proof. Let $\{U_\alpha, x^1_\alpha, \ldots, x^n_\alpha\}$ and $\{V_\beta, y^1_\beta, \ldots, y^n_\beta\}$ be two oriented atlases. If the local orientation of one chart, V_β, coincides with the orientation of the atlas $\{U_\alpha\}$, then both atlases define the same orientation on M. Indeed, it is sufficient to determine the sign of the

Jacobian $\frac{\partial(x_\alpha^1, \ldots, x_\alpha^n)}{\partial(y_\gamma^1, \ldots, y_\gamma^n)}$ at any point $P \in U_\alpha \cap V_\gamma$. Let $P_0 \in U_{\alpha_0} \cap V_\beta$ be an arbitrary point. Connect P and P_0 by a continuous curve φ: $[0, 1] \to M$, $\varphi(0) = P_0$, $\varphi(1) = P$. Consider the infimum t_0 of those numbers $t \in [0, 1]$ for which the sign of the Jacobian at the point $\varphi(t)$ is negative for a certain choice of the charts containing $\varphi(t)$. Then, on the one hand, $t_0 \neq 0$, i.e. $\varphi(t_0) \neq P_0$. On the other hand, the point $\varphi(t_0)$ and the point $\varphi(t)$, at which the Jacobian is negative, lie in the same chart. The choice of the charts $U_\alpha \ni \varphi(t_0)$, $\varphi(t)$, and $V_\beta \ni \varphi(t_0)$, $\varphi(t)$, is insignificant, since the Jacobian is negative at the point $\varphi(t)$ for any two charts U_α and V_γ that contain $\varphi(t)$. Since the Jacobian is a continuous, nonzero function, its sign is negative at the point $\varphi(t_0)$ as well as at the point $\varphi(t')$ for $t' < t_0$. Hence, t_0 is not the infimum. The contradiction shows that both atlases define the same orientation on M. In a similar way one can verify that all charts with a local orientation opposite to that of the atlas $\{U_\alpha\}$ form an oriented atlas. The proposition is proved.

There is another concept of orientation of M. We say that the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in \mathbf{R}^n$ defines the orientation of a Euclidean space. Two bases are assumed to have the *same orientation* if the determinant of the transition matrix is positive. Then $\mathbf{R}^n$ acquires exactly two distinct orientations. To define the orientation of M means to define the orientation of the tangent space T_PM at each point $P \in M$. Let φ: $[0, 1] \to M$ be an arbitrary path. For any basis we can construct at the initial point $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in T_{\varphi(0)} M$ a continuous family of bases $(\mathbf{e}_1(t), \ldots, \mathbf{e}_n(t)) \in T_{\varphi(t)} M$ which begins with $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. All the more, any two families define the same orientation in $T_{\varphi(t)} M$. If M is orientable, the orientation at the terminal point $\varphi(1)$ does not depend on the choice of the path φ. Thus, to define an orientation on M means to define the orientation of the tangent spaces $T_P M$ in such a way that the orientations are consistent along any continuous path connecting the initial and terminal points.

Theorem 1. *If a manifold with boundary M is orientable, the boundary ∂M is also an orientable manifold.*

Proof. Let $\{U_\alpha (x_\alpha^1, \ldots, x_\alpha^n)\}$ be an atlas on M. The last coordinate x_α^n in any chart U_α is a non-negative function. M is orientable by hypothesis, then by replacing, if necessary, in each chart U_α the coordinate x_α^1 by $-x_\alpha^1$, we may assume that the transformation Jacobian is positive at each intersection $U_{\alpha\beta} = U_\alpha \cap U_\beta$. As charts on ∂M we take, as before, the intersections $W_\alpha = U_\alpha \cap \partial M$, and as local coordinates we take the first $(n-1)$ coordinates. Let us demonstrate that $\det \left\| \frac{\partial y_\alpha^i}{\partial y_\beta^j} \right\|_{i,j=1}^n > 0$ at an arbitrary point $P \in W_\alpha \cap W_\beta$. Indeed, in the ambient manifold M the Jacobi

matrix of the coordinate transformation at the point P is of the form

$$\frac{\partial(x_\alpha)}{\partial(x_\beta)}=\begin{pmatrix} \frac{\partial x_\alpha^1}{\partial x_\beta^1} & \dots & \frac{\partial x_\alpha^{n-1}}{\partial x_\beta^1} & \frac{\partial x_\alpha^n}{\partial x_\beta^1} \\ \vdots & & \vdots & \vdots \\ \frac{\partial x_\alpha^1}{\partial x_\beta^{n-1}} & \dots & \frac{\partial x_\alpha^{n-1}}{\partial x_\beta^{n-1}} & \frac{\partial x_\alpha^n}{\partial x_\beta^{n-1}} \\ \frac{\partial x_\alpha^1}{\partial x_\beta^n} & \dots & \frac{\partial x_\alpha^{n-1}}{\partial x_\beta^n} & \frac{\partial x_\alpha^n}{\partial x_\beta^n} \end{pmatrix}.$$

Since $x_\alpha^n = x_\beta^n = 0$ in the intersection $W_\alpha \cap W_\beta$, we have $\frac{\partial x_\alpha^n}{\partial x_\beta^i} = 0$, $1 \leqslant i \leqslant n-1$, so that $\det \frac{\partial(x_\alpha)}{\partial(x_\beta)} = \det \frac{\partial(y_\alpha)}{\partial(y_\beta)} \cdot \frac{\partial x_\alpha^n}{\partial x_\beta^n} > 0$. Hence, $\frac{\partial x_\alpha^n}{\partial x_\beta^n} \neq 0$. Since $x_\alpha^n \geqslant 0$ in the intersection $U_\alpha \cap U_\beta$, we have $\frac{\partial x_\alpha^n}{\partial x_\beta^n}(P) > 0$. Thus, $\det \frac{\partial(y_\alpha)}{\partial(y_\beta)}(P) > 0$. The theorem is proved.

Definition. Let M be an oriented manifold with boundary and let $\{U_\alpha(x_\alpha^1, \dots, x_\alpha^n)\}$, $x_\alpha^n \geqslant 0$, be an atlas defining the orientation on M. The orientation on ∂M defined by the atlas $W_\alpha = U_\alpha \cap \partial M$, $(y_\alpha^1, \dots, y_\alpha^{n-1}) = (x_\alpha^1, \dots, x_\alpha^{n-1})$, is called an *orientation consistent with the orientation on* M.

Examples

4. Consider again a ball $D^2 \subset \mathbf{R}^2$. To define an orientation on D^2, it is sufficient to fix at a point P a basis of the tangent space. If P is chosen on $\partial D^2 = S^1$, the basis can be so chosen that the first vector $\mathbf{e}_1$ is tangent to the boundary and the second vector $\mathbf{e}_2$ points to the centre of the ball (see Fig. 4.84). Then the orientation of the boundary defined by $\mathbf{e}_1$ is consistent with the orientation of D^2. The figure shows another direction of the orientation of a two-dimensional manifold as the direction of the rotation of the tangent plane along the smaller arc from $\mathbf{e}_1$ to $\mathbf{e}_2$.

5. Consider a three-dimensional ball D^3 and its boundary S^2. Figure 4.85 shows how the orientations of D^3 and S^2 can be made consistent using the basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ and the right-hand screw rule.

6. This is an example of a non-orientable manifold. Consider a Möbius band in the form of a square with two parallel sides identified in opposite directions (Fig. 4.86). Fix a basis $(\mathbf{e}_1, \mathbf{e}_2)$ at the point A and construct a continuous family of bases along the curve ABA. At the end of the curve the basis changes the orientation,

i.e. the Möbius band is non-orientable. The boundary is, however, an orientable manifold.

Let us now turn to the description of two-dimensional manifolds. The simplest one is S^2, which is diffeomorphic to $\mathbf{CP}^1$. We have

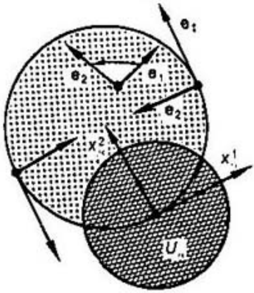

Figure 4.84

already considered the torus $T^2 = S^1 \times S^1$ which also admits the following representation. Let $\mathfrak{G} = \mathbf{Z}(a) \oplus \mathbf{Z}(b)$ be an Abelian group whose generators a and b are defined on $\mathbf{R}^2$ by the shifts $a(x, y) = (x+1, y)$ and $b(x, y) = (x, y+1)$. Considering the factor

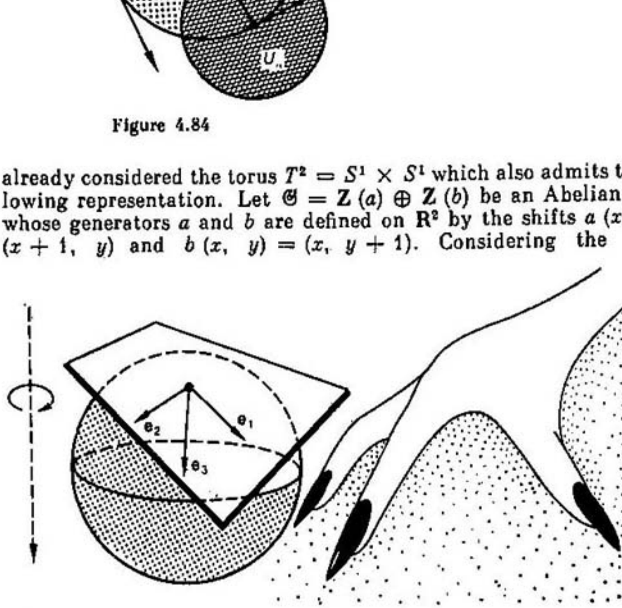

Figure 4.85

space $\mathbf{R}^2/\mathbf{Z} \oplus \mathbf{Z}$, we obtain T^2. The square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ is called the fundamental domain which acts as an elementary "brick": applying the elements of $\mathfrak{G}$ to this domain, we can "pave" the entire plane (see Fig. 4.87). The functions $f(x, y)$ on $\mathbf{R}^2(x, y)$ which are

invariant relative to the transformations $\mathbf{Z}(a) \oplus \mathbf{Z}(b)$ are smooth functions on T^2.

Let us consider on $\mathbf{R}^2(x, y)$ a class of smooth functions f which are invariant relative to another group $\mathfrak{G}(a, b)$ whose generators are defined on $\mathbf{R}^2$ by $b(x, y) = (x, y + 1)$, $a(x, y) = (1 - x, y + 1)$.

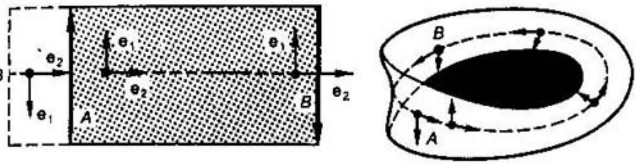

Figure 4.86

(Find relations between a and b!) The fundamental domain is shown in Fig. 4.88. Arrows indicate identification of the sides in accordance with the action of the group. The surface which arises after factorization is shown in Fig. 4.88; it is called a *Klein bottle*. A torus and a Klein bottle are smooth manifolds; a torus is an orientable manifold and a Klein bottle is a non-orientable manifold (verify!).

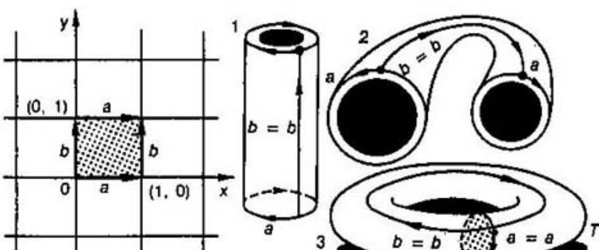

Figure 4.87

These two examples suggest a simple way of constructing other two-dimensional manifolds. Consider gluing on the boundary of a square shown in Fig. 4.89. Since a square is homeomorphic to a disk, we obtain a projective plane. Let us depict this plane in $\mathbf{R}^3$ making appropriate gluings. Figure 4.90 illustrates the process of gluing leading to a model of $\mathbf{RP}^2$ in $\mathbf{R}^3$. We obtain an object which is not an immersed submanifold, two singular points, A and B, being an obstacle. The figure also shows sections of the model by the

planes orthogonal to the singular segment. Although we have failed to immerse $\mathbf{RP}^2$ in $\mathbf{R}^3$ by this method, an immersion does exist (this topic is discussed below).

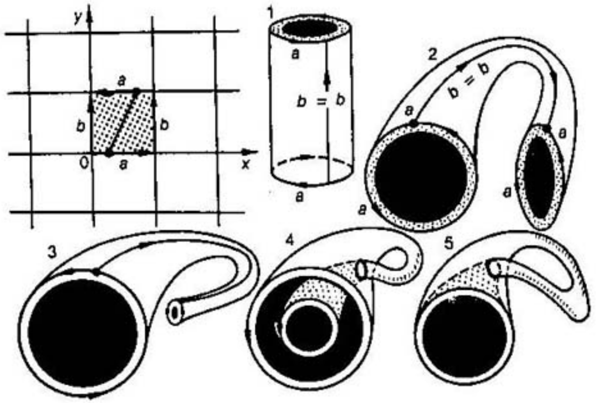

Figure 4.88

Lemma 1. $\mathbf{RP}^2$ *is homeomorphic to gluing along the common boundary of a disk and a Möbius band.*

Figure 4.91 illustrates the *proof* of the lemma.

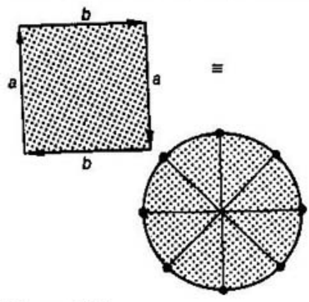

Figure 4.89

An equivalent formulation of Lemma 1 is as follows: elimination of a disk from $\mathbf{RP}^2$ gives a Möbius band. The situation is shown in Fig. 4.92 where the remaining Möbius band occupied such a posi-

tion in $\mathbf{R}^3$ that its boundary circle lies in a two-dimensional plane (but at the expense of two singular points, A and B); such a position of the Möbius band μ_2 in $\mathbf{R}^3$ is called a cross cap.

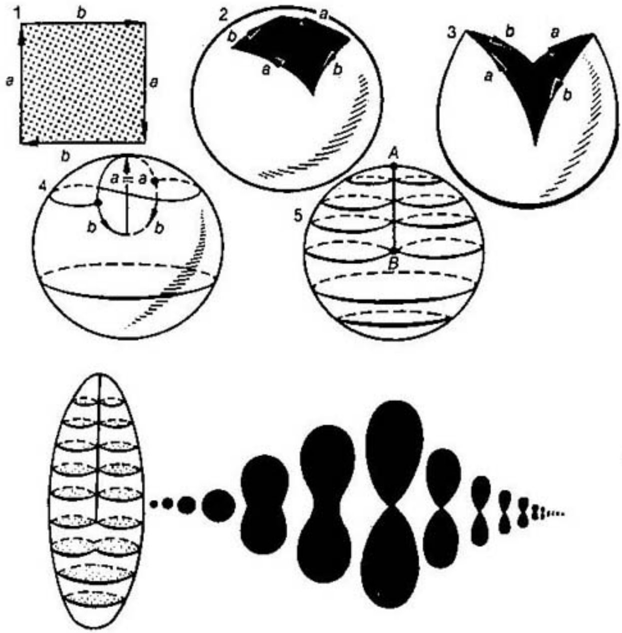

Figure 4.90. Evolution of the sections.

Let us consider T^2 and $\mathbf{RP}^2$ from another point of view. Apply to these manifolds a homeomorphism shown in Fig. 4.93. We may assume that a torus is obtained from S^2 by eliminating two disks and gluing instead of them a handle (see Fig. 4.93) homeomorphic to $S^1 \times D^1$. To obtain an $\mathbf{RP}^2$, one should eliminate from S^2 one disk and glue a Möbius band (see Fig. 4.93). Such an operation is called "gluing a Möbius film". To model this operation in $\mathbf{R}^3$, it is convenient to use a cross cap (see above). Let us apply the above operations, gluing a handle and a Möbius film, two times. What surfaces do we obtain in this case? Consider an octagon (Fig. 4.94) whose sides are marked with letters a, b, c, and d in such a way that moving

clockwise round the octagon, we obtain the word $W = aba^{-1}b^{-1}cdc^{-1}d^{-1}$. Perform gluings, as shown in the figure. The two-dimensional manifold thus obtained is called a pretzel; it is homeo-

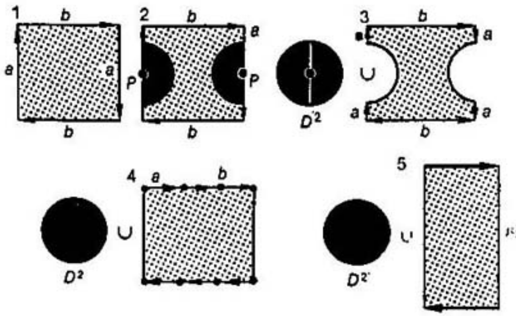

Figure 4.91

morphic to a sphere with two handles: $S^2 + 2r$ (r conventionally denotes a handle).

Lemma 2. *A Klein bottle is homeomorphic to a sphere glued with two Möbius films.*

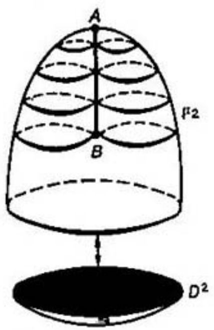

Figure 4.92

The *proof* is illustrated in Fig. 4.95. Gluing two Möbius bands along their common boundary is equivalent to gluing these bands into a sphere with two holes (gluing a cylinder with two Möbius films).

Let us find an immersion of $\mathbf{RP}^2$ in $\mathbf{R}^3$. Consider the immersion of a Möbius band in $\mathbf{R}^3$ as a half of a Klein bottle (see Fig. 4.96). The fact is that cutting the bottle, as is shown in the figure, we obtain two Möbius bands (see Lemma 2). According to Lemma 1, in order to obtain $\mathbf{RP}^2$, it is sufficient to glue the Möbius band with a disk. Let us try to realize this gluing in $\mathbf{R}^3$ so as to obtain the immersion

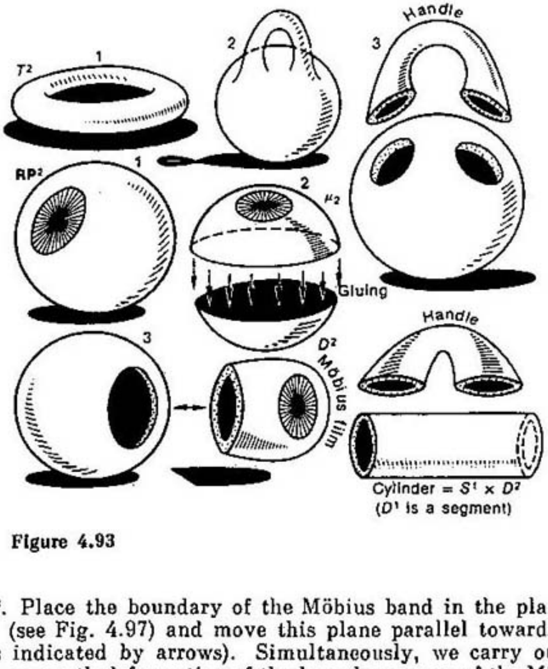

Figure 4.93

$\mathbf{RP}^2 \to \mathbf{R}^3$. Place the boundary of the Möbius band in the plane of the figure (see Fig. 4.97) and move this plane parallel towards the reader (as indicated by arrows). Simultaneously, we carry out in this plane a smooth deformation of the boundary curve of the Möbius band, completing the initial immersion of the Möbius band with the plane swept out by the deformed circle. At the last moment, the circle becomes standardly embedded, and we glue it with a disk to obtain the immersion $\mathbf{RP}^2 \to \mathbf{R}^3$ in question. Figure 4.97 illustrates the structure of the set of self-intersection points.

We have obtained two ways of constructing two-dimensional manifolds (Fig. 4.98):

(1) k handles should be glued to S^2,

(2) s Möbius films should be glued to S^2.

Thus, we get two series of manifolds: $M^2_{g=k} = S^2 + kr$ and $M^2_{\mu=s} = S^2 + s\mu_2$ where M^2_g are orientable and M^2_μ are non-orientable. On M^2_μ there always exist smooth paths lying on the Möbius film, i.e. modifying the orientation of the frames moving along these paths (see above). A priori, there also may exist manifolds of "mixed type", i.e. manifolds obtained from S^2 by gluing k handles and s Möbius films. This operation does not however lead to new manifolds. Let us

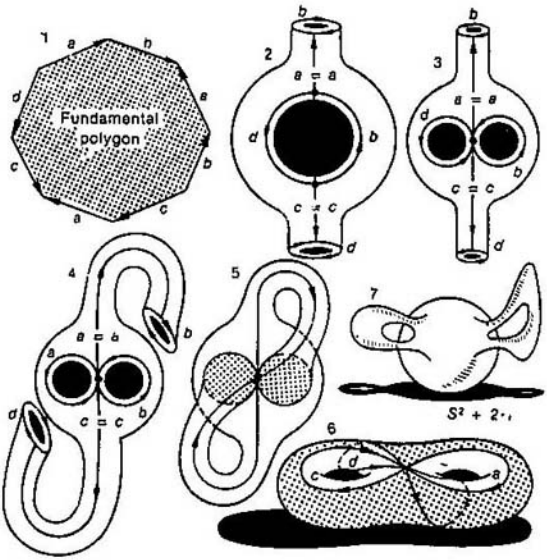

Figure 4.94

consider $S^2 + r + \mu$. Fix one base of handle r and begin to move the second base towards the Möbius film, then move the base along the axis of the Möbius film and leave the film, coming back to the initial position. The handle will find itself in a new position (see Fig. 4.99); this position is called "handle turned inside out" and differs from the ordinary position by the way the bases are glued to S^2. But (see Fig. 4.100) a sphere with a "handle turned inside out" is homeomorphic to a Klein bottle, i.e. (see Lemma 2) to a sphere with two Möbius films. This means that in the presence of a Möbius film a handle turns into two Möbius films; in particular, $S^2 + r + \mu \cong S^2 + 3\mu$, whence $S^2 + kr + s\mu \cong S^2 + (2k + s)\,\mu$, i.e. mani-

folds of "mixed type" are homeomorphic to manifolds of the "non-orientable series".

As was shown, all M_g^2 can be embedded smoothly in $\mathbf{R}^3$. We now prove that manifolds of the non-orientable series can be immersed

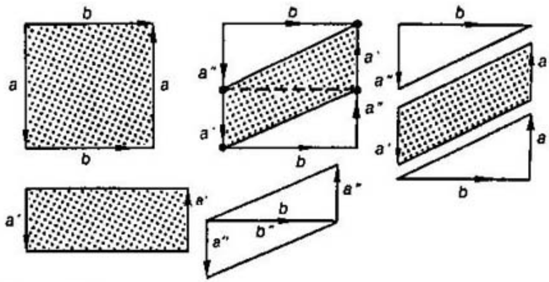

Figure 4.95

smoothly in $\mathbf{R}^3$. This statement has already been proved for $\mathbf{RP}^2 = S^2 + \mu$. Represent M_μ^2 as in Fig. 4.101. Clearly, M_μ^2 is the gluing of several copies of $\mathbf{RP}^2$. Immersing each copy independently in $\mathbf{R}^3$, we obtain the immersion of M_μ^2 in $\mathbf{R}^3$.

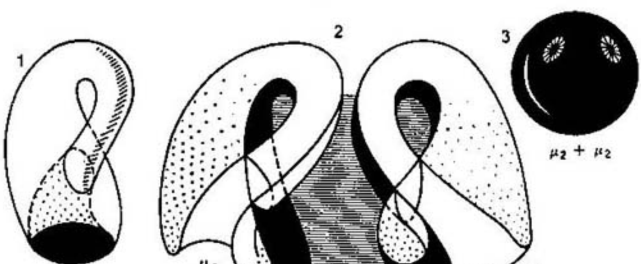

Figure 4.96

It appears that the two series mentioned above include all compact, smooth, closed two-dimensional manifolds.

Theorem 2 (classification theorem). *Any smooth, compact, connected, closed two-dimensional manifold is homeomorphic either to a sphere S^2 with k handles or to a sphere S^2 with s Möbius films.*

It is convenient to *prove* the theorem in several stages. We say that M^2 (of the type stated in Theorem 2) is provided with a triangulation (or is triangulated) if on this manifold there are fixed finitely many points $P_1, \ldots, P_N$ connected (in some order) by finitely many segments of smooth curves γ_α in such a way that:

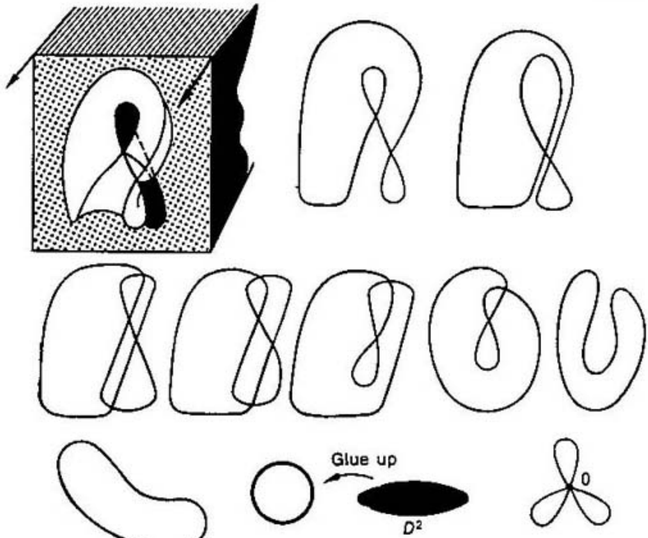

Figure 4.97

(1) each segment γ, starts at a vertex P_α and terminates at a vertex P_β, $P_\alpha \neq P_\beta$, and this segment contains no P_γ other than P_α and P_β,

(2) the set of all these segments subdivides M^2 into finitely many closed triangles with their vertices at points belonging to $\{P_\alpha\}$,

(3) any two triangles Δ_1 and Δ_2 due to this subdivision either do not intersect, or intersect at one common vertex, or intersect along one common side (i.e. along one of the segments γ_i).

These conditions forbid situations shown, for instance, in Fig. 4.102. An example of a triangulation is presented in Fig. 4.103; this is the triangulation of $\mathbf{RP}^2$ consisting of 24 triangles and 13 vertices.

Problem. Construct triangulations of a torus and a Klein bottle. The same M^2 admits infinitely many triangulations.

Lemma 3. *Any two-dimensional smooth, compact, connected, closed manifold M^2 admits a finite triangulation.*

This lemma will be *proved* in Chapter 5.

Let us consider the triangulation of M^2 (of the type stated in Lemma 3). Mark each side of each triangle with a letter and assume that

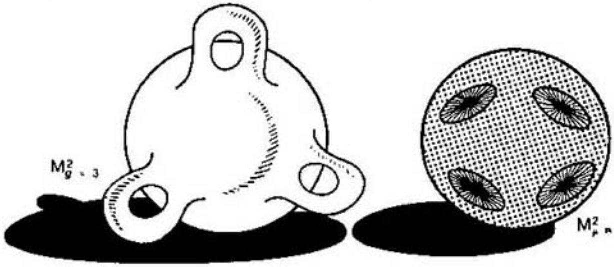

Figure 4.98

different distinct sides are marked with different letters. Provide each side with an arrow indicating the orientation of this side. Directions of the arrows may be chosen arbitrarily. Cut M^2 along all the segments (sides of the triangles). Then M^2 splits into a finite set of triangles each side of which is marked with a letter and provided with an arrow. We assume that while cutting M^2 along any segment, we assign to both borders of the cut the letter that marks the segment (see Fig. 4.104).

Our aim is to glue again all these triangles so as to obtain a polygon. We begin the procedure with choosing a triangle (it is denoted by Δ_1). Consider any side of this triangle; it is endowed with a letter and an orientation. Since each letter appears in the entire set of the sides exactly two times (because M^2 is closed, i.e. it does not have a boundary and each cut has two borders), there can be found another triangle (we denote it by Δ_2) such that one of its sides is marked with the same letter. This triangle is distinct from Δ_1, otherwise the subdivision would not be a triangulation (i.e. one of the cases of Fig. 4.102 would be realized). The triangles Δ_1 and Δ_2 can be glued along this common side in such a way that the arrows point in the same direction. We thus obtain a plane figure whose boundary sides are marked with letters and arrows. Then we choose another letter; again, there can be found a triangle Δ_3 such that one of its sides is marked with the same letter. Glue Δ_3, and so on and so forth. The procedure ends when all the triangles are glued. We do use all the triangles, because if at a certain stage none of the bounadry letters

of the domain thus obtained appeared among the letters of the remaining (not glued) triangles, then by gluing all pairwise letters we would come back to the initial M^2 (gluing eliminates cuts). But by assumption, this M^2 will be disconnected, which contradicts the

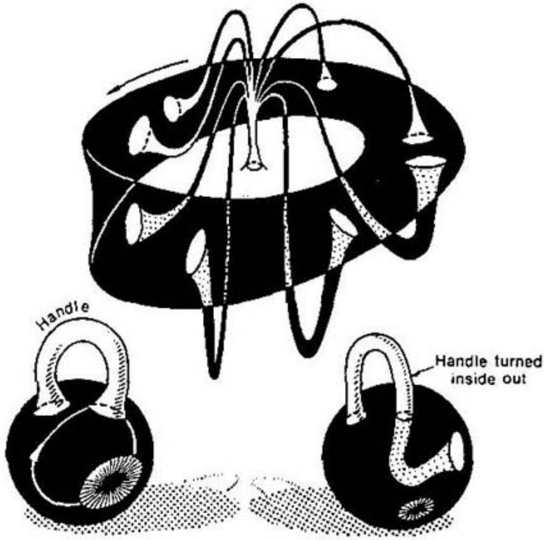

Figure 4.99

condition of the lemma. Thus, we use all the triangles and obtain a plane polygon W. That W is plane follows from the fact that each time we glue a plane triangle to a plane domain only along one side.

The polygon W is not defined uniquely (even for a fixed triangulation ; it has the following properties: (1) W is plane, (2) the boundary ∂W consists of an even number of sides, each being endowed with a letter and an arrow, each letter appears in the boundary two times.

Let us fix an orientation on W and go round the boundary (starting with an arbitrary vertex P), writing successively all the letters. If the direction of motion coincides with the arrow, the power of the corresponding letter is $+1$, otherwise, it is -1. When we come back to the point P we obtain the word $W = a_{i_1} a_{i_2}^{\varepsilon_2} \ldots a_{i_k}^{\varepsilon_k}$,

where $\varepsilon_\alpha = \pm 1$. This word uniquely defines the polygon W (see Fig. 4.105).

With each M^2 (in a certain triangulation) we have associated (not uniquely) a word W. The word W may be considered as a code of M^2,

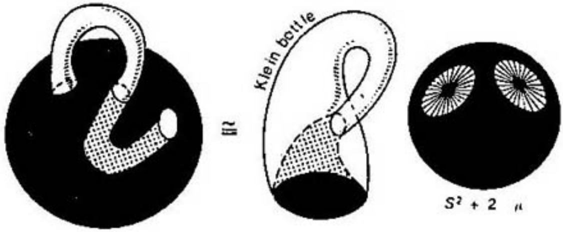

Figure 4.100

and this encoding is not unique: infinitely many codes correspond to a single M^2. Our aim is to prove that any M^2 (of the above type) is homeomorphic either to M_g^2 or to M_μ^2. To this end, we reconstruct

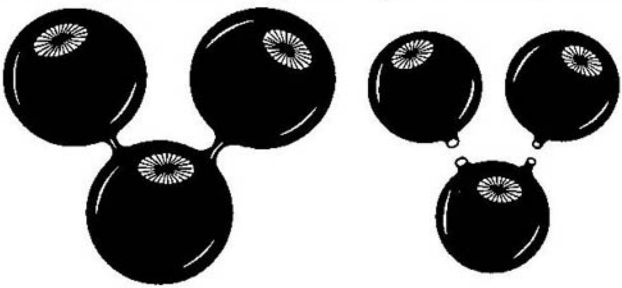

Figure 4.101

the polygon via certain operations appropriate to certain homeomorphisms of the initial M^2 (these operations will not preserve the triangulation, triangulation ceases to be effective as soon as we associate with M^2 its code, the word W). If all the necessary gluings on ∂W are made, we obtain the initial manifold.

Lemma 4. *The word W can be reconstructed (by a homeomorphism of M^2) so that all the vertices of the polygon W are glued into a single point.*

Proof. Let us divide the vertices of W into the classes of equivalent verices. Two vertices are assumed to be equivalent if they are glued

into a single point under identifications on ∂W. If W has only one class of equivalent vertices, the lemma is proved. Let W contain at least two classes of equivalent vertices, $\{P\}$ and $\{Q\}$. We may assume that there exists a side a, such that its beginning belongs to

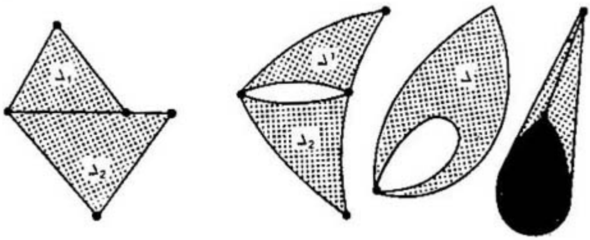

Figure 4.102

$\{P\}$ and the end to $\{Q\}$. Reconstruct W by a homeomorphism of M^2 (see Fig. 4.106). We recall that each letter (in this case c) appears in W exactly two times. As a result of the reconstruction we obtain W' with the following changes in $\{P\}$ and $\{Q\}$: $\{P\}$ has lost one ver-

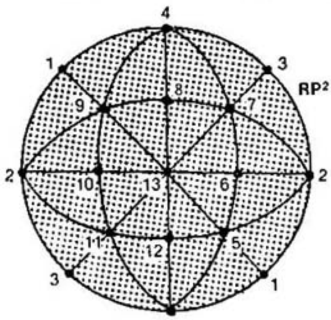

Figure 4.103

tex, and $\{Q\}$ has gained one vertex. Choosing another vertex in $\{P\}$ we can also eliminate it by a certain class (not necessarily by $\{Q\}$). Thus, repeating the procedure, we can reduce the number of vertices in $\{P\}$ until it contains only one vertex. The last stage: since $\{P\}$ has only one vertex, the adjacent sides are of the form a and a^{-1} (see Fig. 4.107). The lemma is proved.

Lemma 5. *Let $W = Abaa^{-1}cB$. Then there exists a homeomorphism of M^2 which transforms W into the word $W' = AbcB$.*

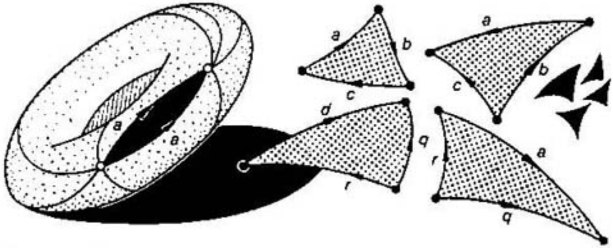

Figure 4.104

Proof. Elimination of aa^{-1} is performed exactly in the same way as the elimination of the last vertex in Lemma 4. The lemma is

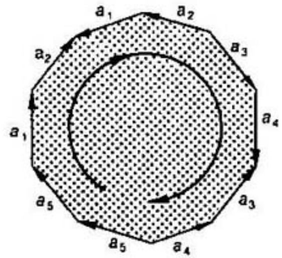

Figure 4.105. $W = a_1a_2a_1^{-1}a_2^{-1}a_3a_4a_3^{-1}a_4^{-1}a_5a_5$

proved. Here A and B denote those parts of W that remain unchanged under the reconstruction.

Lemma 6. *If* $W = BaAaC$, *there exists a homeomorphism of M^2 such that W turns into $W' = BA^{-1}aaC$.*

The *proof* of the lemma is illustrated in Fig. 4.108.

Lemma 7. *If $W = AaRbPa^{-1}Qb^{-1}B$, there exists a homeomorphism of M^2 which transforms W into $W' = AQPaba^{-1}b^{-1}RB$.*

The *proof* is illustrated in Fig. 4.109.

Lemma 8. *If W (subjected to all the preceding operations) contains a pair of letters a and a^{-1}, i.e. $W = AaBa^{-1}C$, $B \neq \varnothing$, there exists a pair b, b^{-1} such that $W = AaDbQa^{-1}Rb^{-1}T$, where $B = DbQ$, $C = Rb^{-1}T$, i.e. for any pair a, a^{-1} ($B \neq \varnothing$) there exists a coupled pair b, b^{-1}.*

Proof. Assume the converse: let for any $b \in B$ the corresponding letter b^ε ($\varepsilon = \pm 1$) appears in the same word B. Glue W along a. The result is shown in Fig. 4.110. Thus, all the vertices of W are

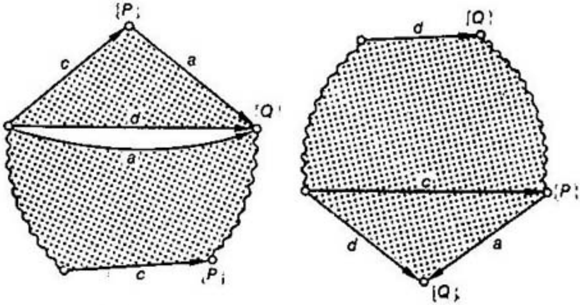

Figure 4.106

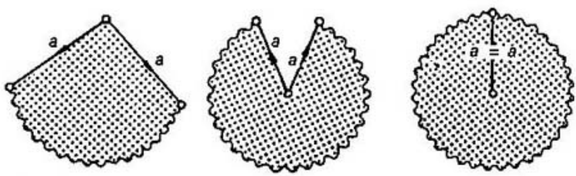

Figure 4.107

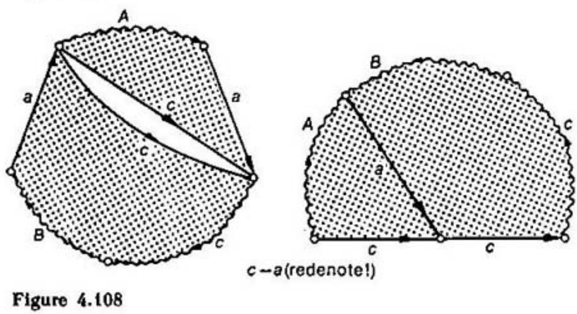

Figure 4.108

divided into at least two classes of non-equivalent vertices. Since the operation of Lemma 4 has been applied to W, we come to contradiction. It remains to prove that $\varepsilon = -1$. If ε were equal to $+1$, we

would obtain contradiction with the form of W implied by Lemma 6. The lemma is proved.

Remark. These lemmas have a feature that each lemma preserves the property of W proved by the preceding lemmas.

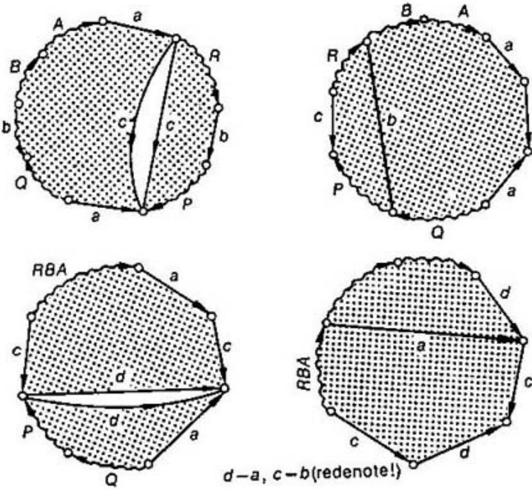

Figure 4.109

Thus, W is represented as the product of the expressions $aba^{-1}b^{-1}$ (*commutators*) and cc (*squares*). This follows from the fact that the expressions $—a—a—$ are reduced to $—aa—$ and the expressions

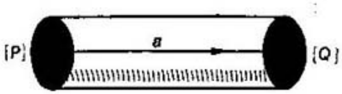

Figure 4.110

$—a—a^{-1}—$ to $—aba^{-1}b^{-1}—$. It remains to analyse the case where the word includes both commutators and squares.

Lemma 9. *If* $W = Aaba^{-1}b^{-1}BccQ$, *there exists a homeomorphism which transforms* W *into* $W' = Mp^2q^2d^2N = Aabd^{-1}B^{-1}bad^{-1}Q$.

The *proof* is illustrated in Fig. 4.111. This operation turns W into the word $\widetilde{W} = Aabd^{-1}B^{-1}bad^{-1}Q$. It remains to "collect to-

gether" the three squares, and this can be done by Lemma 6. The lemma is proved.

This lemma is a formal proof of the assertion that manifolds of the "mixed type" are homeomorphic to non-orientable manifolds of

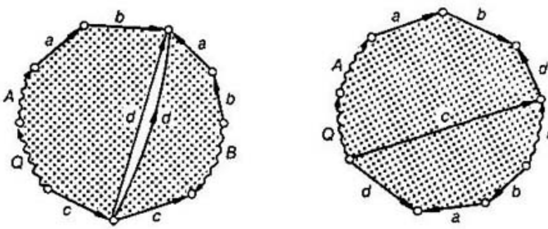

Figure 4.111

the second series (see above). Combining the lemmas proved above, we obtain

Lemma 10. *Let W be one of the codes of M^2. Then there exists a homeomorphism of M^2 which turns W into one of the following words:*

(1) $W = aa^{-1}$,

(2) $W = a_1b_1a_1^{-1}b_1^{-1} \dots a_gb_ga_g^{-1}b_g^{-1}$,

(3) $W = c_1c_1c_2c_2 \dots c_kc_k$.

Which manifolds correspond to these three types? Clearly, in case (1) M^2 is homeomorphic to a sphere (Fig. 4.112). In case (2) M^2

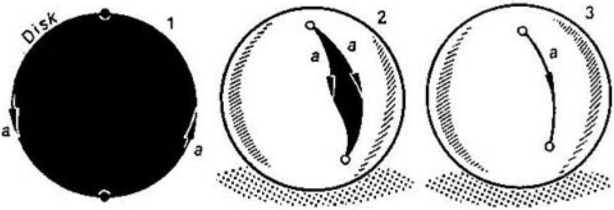

Figure 4.112. $W = aa^{-1}$ is not a triangulation.

is represented as S^2 with g handles. The number g is called the *genus of a surface.* Indeed, for $g = 1$ the word $W = a_1b_1a_1^{-1}b_1^{-1}$ defines a torus (see above); for $g = 2$ the word W' $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}$ defines a pretzel (see above); and so on by induction. In case (3) M^2 is homeomorphic to a sphere with k Möbius films. Indeed, if

$k = 1$, $W = c_1c_1$ defines $\mathbf{RP}^2$ (see above); if $k = 2$, $W = c_1c_1c_2c_2$ defines a Klein bottle, etc. Thus, Theorem 2 is proved.

There exist some other convenient representations of M^2.

Theorem 3. *Any smooth, compact, connected, closed two-dimensional manifold M^2 can be represented in the form $W = a_1a_2 \ldots a_N a_1^{-1} a_2^{-1} \ldots$ $\ldots a_{N-1}^{-1} a_N^{\varepsilon}$ where $\varepsilon = -1$ if and only if $M^2 = M_g^2$ (is orientable),*

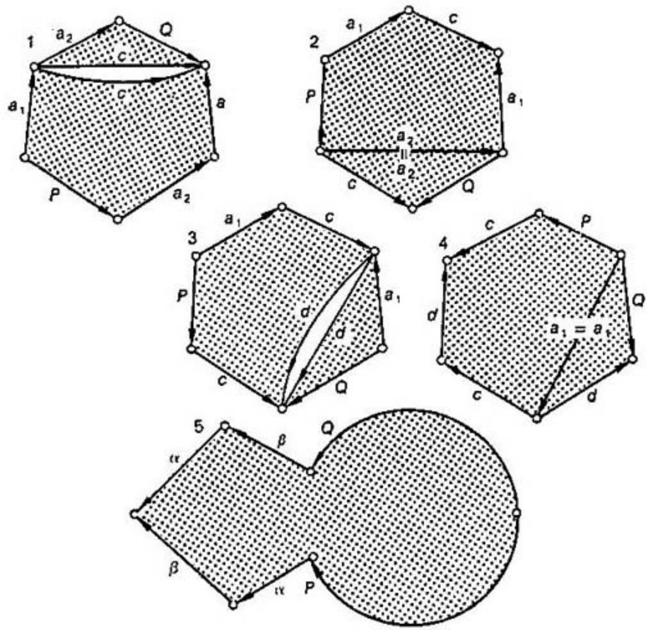

Figure 4.113

in this case N is even; and $\varepsilon = +1$ (N is arbitrary) if and only if $M^2 = M_\mu^2$ (is non-orientable).

Proof. Note that this representation of M^2 is called a canonical symmetric form. That W is reduced (by the elementary operations described above) to words (1) or (2) follows from the previous discussion. Consider, for certainty, $\varepsilon = -1$. We need to prove that the representation described in this theorem is possible for any M_g^2. For what follows see Fig. 4.113. Since $N = 2g$ is even, W is reduced

to $W = a_1b_1a_1^{-1}b_1^{-1}\dots a_gb_ga_g^{-1}b_g^{-1}$, where g is arbitrary, which is what was required. The case $\varepsilon = +1$ is treated similarly (verify!). The theorem is proved.

Here we do not have at our disposal the apparatus for proving that M_g^2 are not homeomorphic to one another for $g_1 \neq g_2$, that M_μ^2 are not homeomorphic for $\mu_1 \neq \mu_2$, and also that M_g^2 and M^2 are not homeomorphic. We report these facts without proof.

4.6. RIEMANNIAN SURFACES OF ALGEBRAIC FUNCTIONS

First, we recall simple properties of complex-valued functions which we shall need below. Let C^n be referred to the coordinates $z^1, \dots, z^n$; if $\gamma(t) = (z^1(t), \dots, z^n(t))$ is a smooth curve, then

$$l(\gamma)_a^b = \int \sqrt{\sum_{k=1}^n \frac{dz^k}{dt}\frac{d\bar{z}^k}{dt}}\, dt = \int_a^b \sqrt{\sum_{k=1}^n \left(\frac{dx^k}{dt}\right)^2 + \left(\frac{dy^k}{dt}\right)^2}\, dt,$$

where $dz^\alpha = dx^k + idy^\alpha$. For $n = 1$ we have

$$\frac{\partial}{\partial x} = \frac{\partial z}{\partial x}\frac{\partial}{\partial z} + \frac{\partial \bar{z}}{\partial x}\frac{\partial}{\partial z} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}},$$

$$\frac{\partial}{\partial y} = \frac{\partial z}{\partial y}\frac{\partial}{\partial z} + \frac{\partial \bar{z}}{\partial y}\frac{\partial}{\partial \bar{z}} = i\left(\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar{z}}\right),$$

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right);$$

$$\frac{\partial z}{\partial \bar{z}} = \frac{\partial \bar{z}}{\partial z} = 0.$$

Let us identify $C^n\{z^k\}$ with $R^{2n}\{x^k, y^k\}$; then, as was already noted, any polynomial $f(z^1, \dots, z^n)$ can be written in the variables $\{x^k, y^k\}$ as $g(x^1, y^1, \dots, x^n, y^n)$. Conversely, any polynomial $g(x^1, y^1, \dots, x^n, y^n)$ can be written in the variables $\{z^k, \bar{z}^k\}$ as $f(z^1, \bar{z}^1, \dots, z^n, \bar{z}^n)$.

Lemma 1. *The polynomial $f(z^1, \bar{z}^1, \dots, z^n, \bar{z}^n)$ does not depend on $\bar{z}^\alpha$ if and only if $\partial f/\partial \bar{z}^\alpha \equiv 0$.*

Proof. The direct statement is obvious. Verify that if $\frac{\partial f}{\partial \bar{z}^\alpha} \equiv 0$, then $f(z^1, \bar{z}^1, \dots, z^n, \bar{z}^n)$ does not contain the variable $\bar{z}^\alpha$. Assume the converse and represent f as a polynomial in the powers of $\bar{z}^\alpha$: $f = \omega(\bar{z}^\alpha)^p + \dots$, where the coefficients $\omega, \dots,$ do not de-

pend on $\bar{z}^{\alpha}$. Let p be the maximal power of $\bar{z}^{\alpha}$, then $\frac{\partial f}{\partial \bar{z}^{\alpha}} = p \times \omega (\bar{z}^{\alpha})^{p-1} + \ldots$, which contradicts the condition. The lemma is proved.

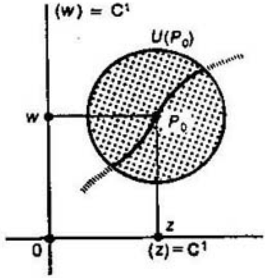

Figure 4.114

We remind that the function $f(z^1, \bar{z}^1, \ldots, z^n, \bar{z}^n)$ is called *analytic* if $\partial f/\partial \bar{z}^{\alpha} = 0$, $1 \leqslant \alpha \leqslant n$. For $n = 1$ we have $\mathbf{R}^2(x, y) = \mathbf{C}^1(z) = \mathbf{R}^2(z, \bar{z})$, $f(x, y) = g(z, \bar{z}) = u(x, y) = iv(x, y)$. If $\frac{\partial f}{\partial \bar{z}} = 0$, then

$$\frac{\partial}{\partial x}(u + iv) + i\frac{\partial}{\partial y}(u + iv) = 0,$$

$$\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) = 0, \quad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$

$$u_{xx} = v_{xy} = -u_{yy}, \; u_{xx} + u_{yy} = 0, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial z\, \partial \bar{z}}.$$

We shall need a complex analogue of the implicit function theorem. The proof of the complex version of the theorem can be found in a course on the theory of functions of a complex variable.

Proposition 1. *Let $f(z, w)$ be a complex-analytic function on $\mathbf{C}^2(z, w)$. Consider the equation $f(z, w) = 0$ in the variable w, and let the relation $\frac{\partial f}{\partial w} \neq 0$ be satisfied at a point $P_0 \in \{f = 0\}$. Then there exists an open neighbourhood $U(P_0)$ in $\mathbf{C}^2$ such that in this neighbourhood the function $w = g(z)$ has the following properties: (1) $g(z)$ is complex-analytic, (2) $w = g(z)$ is a solution of the equation $f(z, w) = 0$ in $U(P_0)$, i.e. $f(z, g(z)) = 0$ in $U(P_0)$, and this solution is unique in $U(P_0)$.*

Geometrically, it is convenient to consider the solution $w = g(z)$ as a "graph" (see Fig. 4.114).

Definition. Let $f(z, w)$ be a polynomial in the variables z, w in $\mathbf{C}^2$ and let the equation $f(z, w) = 0$ be solvable for w in an open neighbourhood $U(P_0)$ (for example, let $\frac{\partial f}{\partial w} \neq 0$). Then the function $w = g(z)$, which is a solution of this equation, is called *algebraic*, and the set of points (z, w) such that $f(z, w) = 0$ (i.e. the zero-level

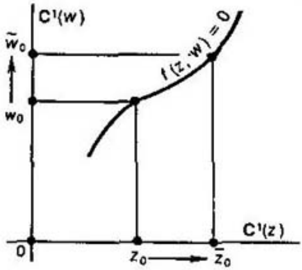

Figure 4.115

surface of $f(z, w)$) is called the *Riemannian surface for the algebraic function* $w = g(z)$ at those points where $w = g(z)$ is defined.

Note that if $\frac{\partial f}{\partial z} \neq 0$ at a point $P_0 = (z_0, w_0)$, then in $U(P_0)$ there exists a solution $z = \varphi(w)$, i.e. the equation $f = 0$ can be solved for z. In this case $\varphi(w)$ satisfies conditions (1) and (2) of Proposition 1. Thus, the condition of local solvability of the equation $f(z, w) = 0$ is of the form grad $f = \left(\frac{\partial f}{\partial z}, \frac{\partial f}{\partial w}\right) \neq 0$, where grad f is the "complex gradient" of the function f.

Let us write the polynomial $f(z, w)$ in a general form, representing it, for example, as a power series of z: $f = a_0(w) z^n + a_1(w) z^{n-1} + \ldots$, where the coefficients $a_k(w)$, $0 \leqslant k \leqslant n$, are polynomials of w. It is worth noting that Riemannian surfaces of algebraic functions are non-compact and "go to infinity" in $\mathbf{C}^2$. (We assume that f is not identically zero.) Indeed, fix a point $w_0 \in \mathbf{C}^1(w)$; this leads to the equation $a_0(w_0)z^n + a_1(w_0) z^{n-1} + \ldots = 0$ for z. According to the familiar algebraic theorem this equation always has a complex root, i.e. there exists a point z_0 such that $f(z_0, w_0) = 0$. Since w_0 can be made to tend to infinity along $\mathbf{C}^1(w)$, the surface $f(z, w) = 0$ also goes to infinity (see Fig. 4.115). Since we deal with a polynomial, we can replace inhomogeneous coordinates in $\mathbf{CP}^2$ by homogeneous coordinates x^1, x^2, x^3, setting $z = x^1/x^3$, $w = x^2/x^3$. Then

$f(z, w) = \Sigma a_{pq} w^p z^q$ is reduced to a homogeneous polynomial $\Sigma a_{pq} (x^1)^q (x^2)^p (x^3)^{s-(p+q)}$, where s is the maximal degree of the monomials $w^p z^q$, $\deg (w^p z^q) = p + q$. Thus, we can "compactify" a Riemannian surface by transferring it from $\mathbf{C}^2$ into $\mathbf{CP}^2$. Since the equation $g(x^1, x^2, x^3) = 0$ is a polynomial on $\mathbf{CP}^2$, this level surface is compact in $\mathbf{CP}^3$. We shall not go into details of this compactification, for the procedure is not trivial and we shall not need it below.

Let us discuss the structure of a Riemannian surface from the topological point of view. From the real point of view, the equation $f(z, w) = 0$ splits into two equations: $\operatorname{Re} f = 0$ and $\operatorname{Im} f = 0$ in $\mathbf{R}^4$, i.e. at the points of "general position" the set $f = 0$ is a two-dimensional (real) surface.

Theorem 1. *Suppose the polynomial $f(z, w)$ is of the form $f = w^q - P_n(z)$ and suppose the polynomial $P_n(z)$ does not have multiple roots. Then the equation $f(z, w) = 0$ defines a smooth two-dimensional (over* $\mathbf{R}$*) submanifold in* $\mathbf{C}^2(z, w)$.

Proof. Let the point P_0 belong to $\{f = 0\}$ and have the coordinates (z_0, w_0), where $w_0 \neq 0$; then $\left.\frac{\partial f}{\partial w}\right|_{P_0} = q w_0^{q-1} \neq 0$ at P_0, i.e. by the implicit function theorem (see Proposition 1), in a neighbourhood of P_0 the surface $\{f = 0\}$ is a smooth two-dimensional submanifold, the graph of the algebraic function $w = g(z)$. In our case the solution $w = g(z)$ is of the form $w = \sqrt[q]{P_n(z)}$. It only remains to consider the points $P_0 = (z_0, 0)$. We assert that $\operatorname{grad} f(P_0) = \left(-\frac{dP_n(z_0)}{dz}, 0\right) \neq 0$. If $\frac{d}{dz} P_n(z_0) = 0$, we have $P_n(z_0) = 0$ and $\frac{d}{dz} P_n(z_0) = 0$ at $(z_0, 0)$, whence z_0 is a multiple root of $P_n(z)$, which contradicts the hypothesis. Thus $\operatorname{grad} f(P_0) \neq 0$, and in an open neighbourhood U of the point $(z_0, 0)$ the surface $\{f = 0\}$ is given by the smooth graph $w = g(z)$, where $g(z)$ is an analytic function. The theorem is proved.

Let us now examine the main properties of algebraic functions.

(1) An algebraic function is usually many-valued, i.e. for some argument z the function $w = g(z)$ has several values: for example, $w = \sqrt[q]{z^n}$, where q and n are relatively prime. The simplest case is the function $w = \sqrt{z}$ which has exactly two values $w = \pm \sqrt{z_0}$ for each $z_0 \neq 0$. The function $z = \alpha(w)$ can also be many-valued. The linear function $w = az + b$ is single-valued. The many-valuedness of $w = g(z)$ can be interpreted as follows. Consider the orthogonal projection $\pi\colon \mathbf{C}^2 \to \mathbf{C}^1(z)$, $\pi(z, w) = (z, 0)$; the complex straight line $H(z_0)$ parallel to $\mathbf{C}^1(w)$ is the inverse image of any point $z_0 \in \mathbf{C}^1(z)$. The Riemannian surface $\Gamma = \{f = 0\}$, considered as the graph of $w = g(z)$, is also projected onto $\mathbf{C}^1(z)$, the inverse image of any point $z_0 \in \mathbf{C}^1(z)$ under $\pi\colon \Gamma \to \mathbf{C}^1(z)$ being represented by

all the values of the function $w = g(z)$ at the point z_0 (Fig. 4.116). Since the function $w = g(z)$ is algebraic (i.e. f is a polynomial), the entire plane $\mathbf{C}^1(z)$ is the image of the surface Γ under the projection $\pi\colon \Gamma \to \mathbf{C}^1(w)$, i.e. for any $z_0 \in \mathbf{C}^1(z)$ there exists at least one solution of the polynomial equation $f(z_0, w) = 0$.

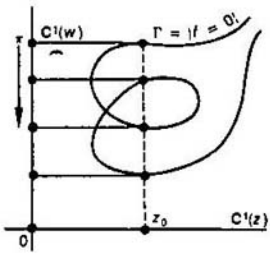

Figure 4.116

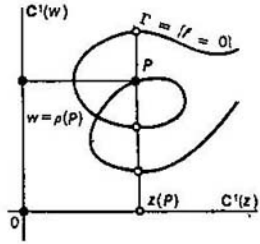

Figure 4.117

Let us introduce a new variable by writing $w = g(z)$ in the form $w = \rho(P)$, there $P \in \Gamma$ is a variable point on the surface Γ (see Fig. 4.117). Apparently, the function $w = \rho(P)$ is single-valued because for each value of the argument $P \in \Gamma$ this function has exactly one value. Thus, we have succeeded in making the algebraic function single-valued (in the new variables), but at the expense of considerable complication of the domain of the argument: in the former case the argument z varied in $\mathbf{C}^1(z)$, and upon substitution the new argument P varies on the two-dimensional manifold Γ. Hence, the Riemannian surface $\Gamma = \{f = 0\}$ of an algebraic function $w = g(z)$ is the domain where the function is single-valued. This property is sometimes used to give another equivalent definition of a Riemannian surface.

(2) Since for each $z_0 \in \mathbf{C}^1(z)$ there exist, in general, many values $w_i = g(z_0)$ (their number is equal to the number of the roots of the polynomial equation $f(z_0, w) = 0$)), one can define in an open neighbourhood of each $z_0 \in \mathbf{C}^1(z)$ a set of continuous (even smooth) functions $w = \{\varphi_i(z)\}$, $1 \leqslant i \leqslant k$, where k is the degree of the polynomial $a_0(z)\,w^k + a_1(z)\,w^{k-1} + \ldots + a_k(z) = f(z, w)$ in the variable w. Each of these functions describes the variation of a certain root of the equation $f(z, w) = 0$ as z changes. The function $\varphi_i(z)$, $1 \leqslant i \leqslant k$, can be continued to all values of the variable z, and at certain points these continued functions can coincide or even interchange. The functions $\varphi_i(z)$ are called *branches of the algebraic function* $w = g(z)$. Each branch describes the behaviour of a

certain root of the equation $f(z, w) = 0$; in this equation z is considered as a parameter and w as the quantity to be sought. The values of the function $w = g(z)$ that are the roots of the equation $f(z_0, w) = a_0(z_0) w^k + a_1(z_0) w^{k-1} + \ldots + a_k(z_0) = 0$ "hang" over each $z_0 \in \mathbf{C}^1(z)$. It follows from the implicit function theorem that each branch defines (almost at all ponts) a smooth, even complex-analytic, submanifold in $\mathbf{C}^2$, since $\varphi_i(z)$ (locally) a complex-analytic function.

(3) Let us associate with each point $z_0 \in \mathbf{C}^1(z)$ a number $k(z_0)$ equal to the number of distinct roots of the equation $f(z_0, w) = 0$.

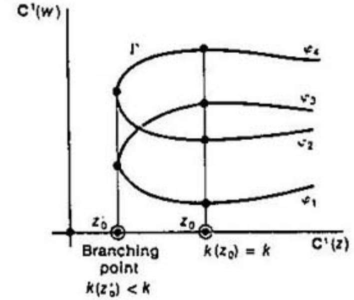

Figure 4.118

Apparently, $k(z_0) \leqslant k$; If all the roots of $f(z_0, w) = 0$ are simple, then $k(z_0) = k$; if the equation has multiple roots, then $k(z_0) < k$. The number $k(z_0)$ is equal to the number of distinct values of $w = g(z)$ at the point z_0. The points z_0, where $k(z_0) < k$, also have the property that at them certain branches $\varphi_i(z)$ of the function $w = g(z)$ merge, thereby reducing the number of distinct values $\{\varphi_i(z)\}$ over z_0 (see Fig. 4.118). Suppose $f(z, w) = w^q - P_n(z)$, and the polynomial $P_n(z)$ does not have multiple roots. According to Theorem 1, the surface $\Gamma = \{f = 0\}$ is a complex-analytic submanifold in $\mathbf{C}^2$. The points $z_0 \in \mathbf{C}^1(z)$ for which $k(z_0) < k$ are called *branching points of an algebraic function* $w = g(z)$ (this term is discussed below). In the example $w = \sqrt[q]{P_n(z)}$, the branching points (in the finite part of $\mathbf{C}^1(z)$, i.e. in the plane $\mathbf{C}^1(z)$ not completed with "infinity") are represented by the roots of the polynomial $P_n(z)$; and if $z_1, \ldots, z_n$ are the roots of $P_n(z)$ (they are all simple), then $k(z_\alpha) < k = q = \deg_w(w^q - P_n(z))$, since $k(z_\alpha) = 1$, $1 \leqslant \alpha \leqslant n$. The "infinity" point is not considered for the time being. Thus, all branching points of the function $w = \sqrt[q]{P_n}(z)$ are isolated.

(4) The branching points (in the finite part of $C^1(z)$) have the property that several branches of the function $w = g(z)$ merge at these points. Since a branching point is isolated, it may be considered as the centre of a sufficiently small disk D^2 which does not contain other branching points, so that $k(z) = k$ for $z \in D^2$, $z \neq z_0$. Consider a circle in D^2 with centre at z_0 and move along this circle round z_0. Generally, the branches of the function $g(z)$ interchange, so that making an appropriate number of revolutions round z_0, we can pass from one branch to another.

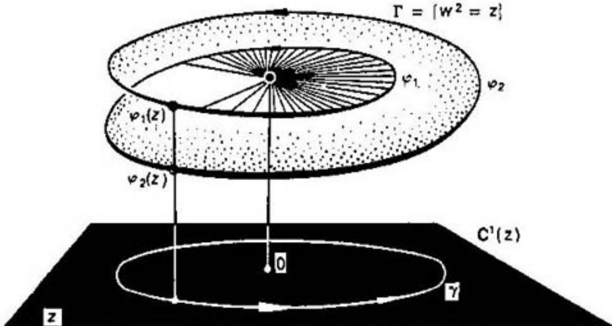

Figure 4.119

Here is an example. Let $f = w^2 - P_n(z)$, where $P_n(z)$ does not have multiple roots; let $n = 1$, i.e. $f = w^2 - z$. Then $w = g(z) = \sqrt{z}$. Obviously, $k = 2$, $w = \{\varphi_1(z), \varphi_2(z)\}$, where $\varphi_1(z_0) = +\sqrt{z_0}$ and $\varphi_2(z_0) = -\sqrt{z_0}$. If $z = r \cdot e^{i\varphi}$, then $\varphi_1(z) = \sqrt{r}\, e^{i\varphi/2}$, $\varphi_2(z) = -\sqrt{r}\, e^{i\varphi/2}$; $\varphi_1(z) \neq \varphi_2(z)$ for $z \neq 0$; $\varphi_1(0) = \varphi_2(0) = 0$. The point $0 \in C^1(z)$ is the only branching point in the finite part of the plane. Both branches are smooth functions. If one moves round the point 0 (r is constant, $0 \leqslant \varphi \leqslant 2\pi$), the branches $\varphi_1(z)$ and $\varphi_2(z)$ interchange:

$$\varphi_1(z) = \sqrt{r}\, e^{\frac{i\varphi}{2}} \to \sqrt{r}\, e^{\frac{i(\varphi+2\pi)}{2}} = -\sqrt{r}\, e^{\frac{i\varphi}{2}} = \varphi_2(z).$$

This interchange is schematically shown in Fig. 4.119. The Riemannian surface itself does not contain singular points, since Γ is a smooth submanifold in C^2.

Let us analyse the global topological properties of a Riemannian surface. We shall confine ourselves to the case $f(z, w) = w^2 - P_n(z)$, where P_n does not have multiple roots. The surface Γ is defined in C^2

as the graph $w = g(z) = \sqrt{P_n(z)}$. Before studying the behaviour of Γ at infinity, let us explain what we mean by "infinity". There are several ways of defining this concept. This is the so-called compactification problem; it is related to a more general question: how can a non-compact, open manifold be transformed into a compact manifold by "adjoining a boundary"? This problem does not, in general, have a universal solution, since the concept of compactification depends on the purpose of a given problem: namely, an open manifold can be compactified in different ways (if this manifold admits compactification).

Let us consider a straight line $\mathbf{C}^1(z)$ and adjoin to it an infinite point, i.e. adjoin this point to the plane $\mathbf{R}^2(x, y)$; in this case $\mathbf{C}^1 \cup \infty$ is transformed into S^2 which will be called a completed complex straight line. There is a canonical way of identifying a completed complex straight line with a one-dimensional space $\mathbf{CP}^1$. Choose for $\mathbf{CP}^1$ the model: $(\lambda z^1, \lambda z^2)$, where $\lambda \neq 0$, and consider the mapping $h: (\lambda z^1, \lambda z^2) \to \frac{\lambda z^1}{\lambda z^2} = \frac{z^1}{z^2} \in S^2 = \mathbf{C}^1(z) \cup \infty$. If $\frac{a}{b} = \frac{c}{d} = \rho$, we may assume that $a = 1$, $c = 1$, i.e. $b = d = \rho^{-1}$ and $(1, \rho^{-1}) = (1, \rho^{-1})$, i.e at "finite" points of the extended plane $\mathbf{R}^2 \cup \infty = S^2$ the mapping h is a homeomorphism; adjoining of the point ∞ preserves this property of h. Thus, we may sometimes assume that $\mathbf{CP}^1 \cong S^2$. Recall that the extended plane can also be identified with S^2 by stereographic projection.

Let us now consider the direct product $S^2 \times S^2$; a complex structure can be introduced in this four-dimensional manifold, since $S^2 \cong \mathbf{CP}^1$, i.e. S^2 is a complex manifold. Fix on the spheres $S^2(z)$ and $S^2(w)$ infinitely distant points ∞_z and ∞_w, and assume that $S^2(z) = \mathbf{C}^1(z) \cup \infty_z$, $S^2(w) = \mathbf{C}^1(w) \cup \infty_w$. Then the point (∞_z, ∞_w) is fixed in $S^2(z) \times S^2(w)$, such that two copies of the sphere, $\infty_z \times S^2(w)$ and $S^2(z) \times \infty_w$, pass through this point. If we eliminate this wedge of two spheres (glued at (∞_z, ∞_w)), we obtain $\mathbf{R}^4$ identified with $\mathbf{C}^2(z, w)$ (see Fig. 4.120). Clearly,

$$(S^2(z) \times S^2(w)) \setminus [(S^2(z) \times \infty_w) \cup (\infty_z \times S^2(w))]$$
$$= (S^2(z) \setminus \infty_z) \times (S^2(w) \setminus \infty_w) = \mathbf{C}^1(z) \times \mathbf{C}^1(w) = \mathbf{C}^2(z, w).$$

This is a particular case of the formula $(X \times Y) \setminus [(X \times y_0) \cup (x_1 \times Y)] \cong (X \setminus x_0) \times (Y \setminus y_0)$, where $x_0 \in X$, $y_0 \in Y$. Conversely, $\mathbf{C}^2(z, w)$ can be transformed from a non-compact, open manifold into a compact manifold by adjoining to the former a boundary, i.e. the wedge of two spheres $(S^2_z \times \infty_w) \vee (\infty_z \times S^2_w)$ glued at the point (∞_z, ∞_w) (by the *wedge of two topological spaces X and Y* we mean the space obtained from X and Y by identifying two points $x_0 \in X$, $y_0 \in Y$; the wedge of X and Y is denoted by $X \vee Y$). This is one of the ways of the compactification of $\mathbf{C}^2$. Another way is to

adjoin to $\mathbf{C}^2$ a point, "infinity", thereby transforming $\mathbf{C}^2$ into a sphere S^4. These two ways of compactification differ significantly; one can prove that $S^2 \times S^2$ and S^4 are not homeomorphic (the proof is beyond the scope of this book).

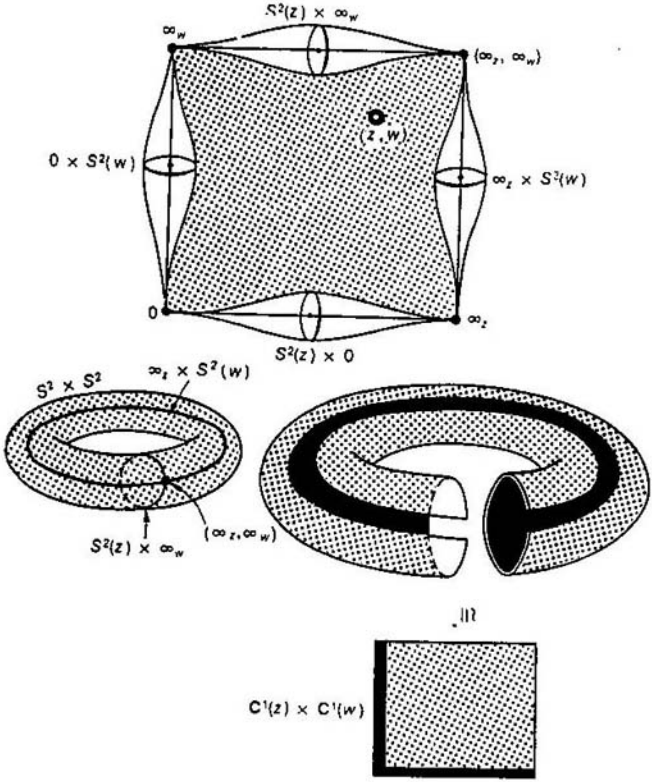

Figure 4.120. $(S^2 \times S^2) \setminus [(S^2(z) \times \infty_w) \cup (\infty_z \times S^2(w))] \cong \mathbf{C}^2(z, w)$.

There is one more method to compactify $\mathbf{C}^2$ and transform it into a compact manifold. We are already familiar (in part) with this method; it lies in introducing projective coordinates. Let us consider $\mathbf{CP}^2$ referred to homogeneous coordinates (x^1, x^2, x^3), then $\mathbf{CP}^2$ is

covered with three charts homeomorphic to $\mathbf{C}^2$:

$$A_1 = \{\lambda(x^1, x^2, x^3),\ x^1 \neq 0\},$$
$$A_2 = \{\lambda(x^1, x^2, x^3),\ x^2 \neq 0\},\ A_3 = \{\lambda(x^1, x^2, x^3),\ x^3 \neq 0\}.$$

Let

$$\alpha_3\colon A_3 \to \mathbf{C}^2(z, w),\ \alpha_3(\lambda(x^1, x^2, x^3)) = \left(\frac{x^1}{x^3}, \frac{x^2}{x^3}\right),$$
$$z = \frac{x^1}{x^3},\ w = \frac{x^2}{x^3},$$

then α_3 is a homeomorphism between the chart A_3 and $\mathbf{C}^2$. The homeomorphisms $\alpha_2\colon A_2 \to \mathbf{C}^2(u, v)$ and $\alpha_1\colon A_1 \to \mathbf{C}^2(\alpha, \beta)$

$$u = \frac{x^1}{x^2},\quad v = \frac{x^3}{x^2},\quad \alpha = \frac{x^2}{x^1},\quad \beta = \frac{x^3}{x^1},$$

are constructed in a similar fashion (see Fig. 4.121). Set

$$\mathbf{CP}_1^1 = S_1^2 = \{\lambda(0, x^2, x^3)\},\quad \mathbf{CP}_2^1 = S_2^2 = \{\lambda(x^1, 0, x^3)\},$$
$$\mathbf{CP}_3^1 = S_3^2 = \{\lambda(x^1, x^2, 0)\},$$

then $\mathbf{CP}^2 = A_1 \cup S_1^2$, $\mathbf{CP}^2 = A_2 \cup S_2^2$, $\mathbf{CP}^2 = A_3 \cup S_3^2$, where the compact two-dimensional manifolds S_1^2, S_2^2, and S_3^2 are homeomorphic

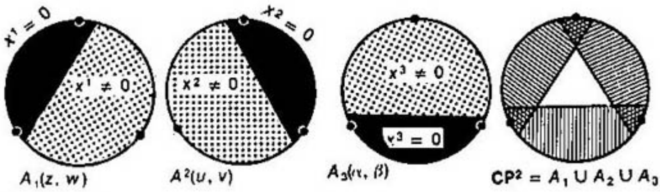

Figure 4.121

to S^2. Thus, $\mathbf{C}^2$ can be compactified by adjoining a sphere S^2, which transforms $\mathbf{C}^2$ into $\mathbf{CP}^2$; this operation can be realized in two ways (see above). Let us consider all the three methods of compactification of $\mathbf{C}^2$ and observe what happens to a Riemannian surface Γ embedded in $\mathbf{C}^2$. Since Γ is non-compact and goes to infinity, it also undergoes compactification, thereby becoming a compact topological space $\widetilde{\Gamma}$. Demonstrate that $\widetilde{\Gamma}$ is a smooth manifold.

Theorem 2. *Let $\mathbf{C}^2$ be compactified to $S^2 \times S^2$ by adjoining to $\mathbf{C}^2$ the wedge of two spheres $S^2 \vee S^2 = (S^2(z) \times \infty_w) \vee (\infty_z \times S^2(w))$. Then the Riemannian surface Γ of the algebraic function $w = \sqrt{P_n(z)}$ (the polynomial P_n has simple roots) is compactified to a compact, smooth, closed two-dimensional manifold $\widetilde{\Gamma}$.*

Proof. We first consider that part of $\widetilde{\Gamma}$ which is contained in $\mathbf{C}^2 = (S^2 \times S^2)\setminus(S^2 \vee S^2)$. According to Theorem 1, this part of $\widetilde{\Gamma}$ is a smooth manifold. Consider $\widetilde{\Gamma} \cap (S^2 \vee S^2)$. We assert that this is a single point, i.e. $\widetilde{\Gamma}$ touches the set of "infinite points" $(S^2(z \times \infty_w) \cup (\infty_z \times S^2(w))$ only at one point. Indeed, the form of $w = \sqrt{P_n(z)}$ implies that the image of "infinity" is represented solely by "infinity", i.e. $\widetilde{\Gamma}$ contains only one infinite point (∞_z, ∞_w) (Fig. 4.122). Let

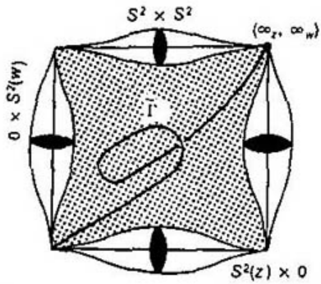

Figure 4.122

us prove that the structure of a smooth manifold can be introduced in a neighbourhood of the point (∞_z, ∞_w) on $\widetilde{\Gamma}$. Define on $S^2 \times S^2$ the coordinates z and w which run $S^2(z)$ and $S^2(w)$. Introduce new variables $u = 1/z$ and $v = 1/w$. Since the transformation $z \to u = 1/z$ is regular on $S^2(z)$, this transformation is also regular on $S^2(z) \times S^2(w)$, so that $\infty_z \to 0_z \in S^2(z)$, $\infty_w \to 0_w \in S^2(w)$. The equation

$$w^2 = P_n(z) = \prod_{k=1}^{n} (z - a_k),$$

where $a_i \neq a_j$ for $i \neq j$, is reduced to

$$\frac{1}{v^2} - \prod_{k=1}^{n} \left(\frac{1}{u} - a_k\right) = 0, \quad \frac{u^n}{v^2} - \prod_{k=1}^{n} (1 - a_k u) = 0,$$

that is

$$v^2 = \frac{u^n}{\prod_{k=1}^{n} (1 - a_k u)}.$$

We now study the structure of zeros of the function $f(u, v) = v^2 - u^n \left(\prod_{k=1}^{n} (1 - a_k u)\right)^{-1}$ in a neighbourhood of the point 0 $(u = 0, v = 0)$. After the change $v^2 = t$, $u = u$ the equation takes the form $t = \frac{u^n}{\prod_{k=1}^{n} (1 - a_k u)}$. Calculation yields $\operatorname{grad} \widetilde{f}(t, u)|_0 = (1, 0)$ for $n > 1$ and $\operatorname{grad} \widetilde{f}(t, u)|_0 = (1, 1)$ for $n = 1$, i.e. $\operatorname{grad} \widetilde{f} \neq 0$ at the point 0, and by the implicit function theorem the surface under study (in the variables (t, u)) is a smooth submanifold in $\mathbf{C}^2$ (in the neighbourhood of 0). With the inverse change, $v = \sqrt{t}$ and $u = u$, the graph of the preceding manifold undergoes square-root transformation. Since the equation $v^2 - t = 0$ defines near the point 0 a smooth submanifold, the initial graph is also a smooth submanifold. The theorem is proved.

Under other compactifications (adjoining a point and adjoining a sphere), the corresponding sets $\widetilde{\Gamma}'$ and $\widetilde{\Gamma}''$ are also smooth submanifolds in the compactified $\mathbf{C}^2$. We shall not prove this statement in a general case, but consider several examples.

Under the compactification $\mathbf{C}^2 \to \mathbf{C}^2 \cup \infty = S^4$, the surface Γ embedded in $\mathbf{C}^2$ is also compactified by a single point, but under $\mathbf{C}^2 \to \mathbf{C}^2 \cup S^2 = \mathbf{CP}^2$ the situation is more complicated. Since $\Gamma \subset A_3 = \{\lambda (x^1, x^2, x^3), x^3 \neq 0\}$, we have to find, under the compactification $\mathbf{C}^2 \to \mathbf{CP}^2$, the intersection of $\widetilde{\Gamma}''$ with the adjoined sphere $S^2 = \mathbf{CP}^1 = S^2_3$. Let $n > 2$. The change $z = \frac{x^1}{x^3}$, $w = \frac{x^2}{x^3}$ transforms the equation $\prod_{k=1}^{n} (z - a_k) = w^2$ into $(x^2)^2 (x^3)^{n-2} - \prod_{k=1}^{n} (x^1 - a_k x^3) = 0$ (verify!). Since S^2_3 is defined in $\mathbf{CP}^2$ by the equation $x^3 = 0$, the intersection $\widetilde{\Gamma}'' \cap S^2_3$ can be found by setting $x^3 = 0$ in the equation of the surface, which gives the solution $(x^1 = 0$, x^2 is arbitrary, $x^3 = 0)$. This means that $\widetilde{\Gamma}'' \cap S^2_3$ consists of a single point with the coordinates $u = \frac{x^1}{x^2} = 0$, $v = \frac{x^3}{x^2} = 0$ in the chart $A_2 = \{\lambda (x^1, x^2, x^3), x^2 \neq 0\}$ (Fig. 4.123). To study the local structure of the set of solutions to the equation $(x^2)^2 (x^3)^{n-2} - \prod_{k=1}^{n} (x^1 - a_k x^3) = 0$ in a neighbourhood of the point $(x^1 = x^3 = 0, x^2 = \lambda)$, one should make the change $u = \frac{x^1}{x^2}$, $v = \frac{x^3}{x^2}$, which leads to the equation $q = v^{n-2} - \prod_{k=1}^{n} (u - a_k v) = 0$. For $n > 2$ the gradient of the

polynomial q is zero at the point $(0, 0)$, and it is not obvious therefore that the surface is a smooth submanifold at this point. Though a new change can be made, which elucidates this fact, we shall not concentrate on the topic and consider only the cases $n=1$ and $n=2$.

For $n=1$ we have $w^2-z=0$, $\left(\frac{x^2}{x^3}\right)^2-\frac{x^1}{x^3}=0$, $(x^2)^2-x^1x^3=0$. Since $\Gamma\subset A_3$, one has to find $\widetilde{\Gamma}''\cap S_3^2$, i.e. to assume $x^3=0$,

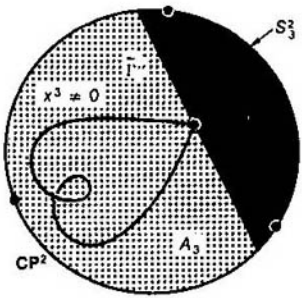

Figure 4.123

which gives: x^1 is arbitrary, $x^2=0$, $x^3=0$. Thus, the intersection of $\widetilde{\Gamma}''$ with infinitely distant sphere S_3^2 is a single point with the coordinates $\alpha=\frac{x^2}{x^1}=0$, $\beta=\frac{x^3}{x^1}=0$ in the chart $A_1=\{\lambda(x^1, x^2, x^3), x^1\neq 0\}$. To examine the neighbourhood of this point, it is convenient to make the change $\alpha=\frac{x^2}{x^1}$, $\beta=\frac{x^3}{x^1}$, which gives $[(x^2)^2-x^1x^3=0]\to\left[\left(\frac{x^2}{x^1}\right)^2-\frac{x^3}{x^1}=0\right]=[\alpha^2-\beta=0]$. Apparently, this equation defines a smooth submanifold in a neighbourhood of the point $\alpha=\beta=0$, and this proves that the compactified surface $\widetilde{\Gamma}''$ is smooth at an infinitely distant point. For $n=2$ we have $w^2-(z-a_1)(z-a_2)=0$ where $a_1\neq a_2$,

$$\left(\frac{x^2}{x^3}\right)^2-\left(\frac{x^1}{x^3}-a_1\right)\left(\frac{x^1}{x^3}-a_2\right)=0,$$
$$(x^2)^2-(x^1-a_1x^3)(x^1-a_2x^3)=0.$$

As for $n=1$, we obtain: if $x^3=0$, then $x^2=\pm x^1$, i.e. the intersection $\widetilde{\Gamma}''\cap S_3^2$ consists of two points $T_1=(x^1=\lambda,\ x^2=\lambda,\ x^3=0)$

and $T_2 = (x^1 = \lambda,\ x^2 = -\lambda,\ x^3 = 0)$. In the chart $A_1 = \{\lambda(x^1, x^2, x^3),\ x^1 \neq 0\}$ we have $T_1 = (\alpha = 1,\ \beta = 0)$, $T_2 = (\alpha = -1,\ \beta = 0)$. Since $x^1 \neq 0$, we may pass to the chart A_1 via the change $\alpha = \frac{x^2}{x^1}$, $\beta = \frac{x^3}{x^1}$, which yields

$$\alpha^2 - (1 - a_1\beta)(1 - a_2\beta) = 0,$$

$$\operatorname{grad}\,[\alpha^2 - (1 - a_1\beta)(1 - a_2\beta)]_{T_1T_2}$$
$$= (2\alpha,\ a_1(1 - a_2\beta) + a_2(1 - a_1\beta))_{T_1T_2} = (\pm 2,\ a_1 + a_2) \neq 0$$

at the points T_1 and T_2, so that the surface $\widetilde{\Gamma}''$ is a smooth submanifold near T_1 and T_2. For $n = 1, 2$ we have proved that the compactification $\Gamma \to \widetilde{\Gamma}''$ which arises under the compactification $\mathbf{C}^2 \to \mathbf{CP}^2$, gives a smooth manifold. This is also true for $n > 2$, but the proof of this fact is beyond the scope of the book.

In what follows we shall assume that an infinitely distant point is adjoint to the surface $\Gamma \subset \mathbf{C}^2$, so that Γ becomes a smooth, compact manifold. Let us study the branching points of the function $w = \sqrt{P_n(z)}$ which is now defined on the entire sphere $S^2 = \mathbf{CP}^1$.

Proposition 2. *Let $w^2 - P_n(z) = 0$, where $P_n(z)$ has only simple roots $a_1, \ldots, a_n$. Then all these roots are the branching points of $w = \sqrt{P_n(z)}$ located in a finite part of the plane, on the sphere S^2 there exists, besides these points, another branching points, infinity, provided the degree of P_n is odd. If the degree of P_n is even, infinity is not a branching point. There are no other branching points for the function $w = \sqrt{P_n(z)}$.*

Proof. Consider the case $w^2 - z = 0$. The function $w = \pm\sqrt{z}$ has two branches: $\varphi_1 = \sqrt{z}$ and $\varphi_2 = -\sqrt{z}$. Consider a circle of finite radius with the centre at 0 and go round this point. The branches φ_1 and φ_2 are interchanged, so that the function cannot be made single-valued and smooth on a sphere. Repeat the procedure, taking the point $\infty \in S^2$ as the centre of the circle. On S^2 all the points are equivalent relative to the transformation group $\frac{az+b}{cz+d}$; in particular, ∞ can be transformed into any "finite" point. A circle with centre at ∞ can also be considered as a circle with centre at 0, so that circumvention round ∞ also interchanges the branches (see Fig. 4.124).

We now turn to a general case. Let a_k be an arbitrary root of $P_n(z)$. Consider a small circle S^1 with centre at a_k which contains no other roots of the polynomial and represent the point $z \in S^1$ in the form $z = a_k + re^{i\varphi}$, where r is the radius of the circle. We

have

$$w(z)=\sqrt{\prod_{p=1}^{n}(z-a_p)}=\sqrt{\prod_{p\neq k}(z-a_p)}\cdot\sqrt{z-a_k}$$
$$=\sqrt{\prod_{p\neq k}(z-a_p)}\cdot r^{\frac{1}{2}}\cdot e^{\frac{i\varphi}{2}}$$

(see Fig. 4.125). Furthermore,

$$w(z_0)=r_0^{\frac{1}{2}}\cdot e^{\frac{i\varphi_0}{2}}\sqrt{\prod_{p\neq k}(z-a_p)}\to r_0^{\frac{1}{2}}e^{\frac{i(\varphi_0+2\pi)}{2}}\cdot\sqrt{\prod_{p\neq k}(z-a_p)}$$

under circumvention round a_k. Since this circumvention causes the argument of $(z-a_k)$ to change by 2π, $\sqrt{z-a_k}$ changes the sign. At the same time, circumvention round a_k does not increase the argu-

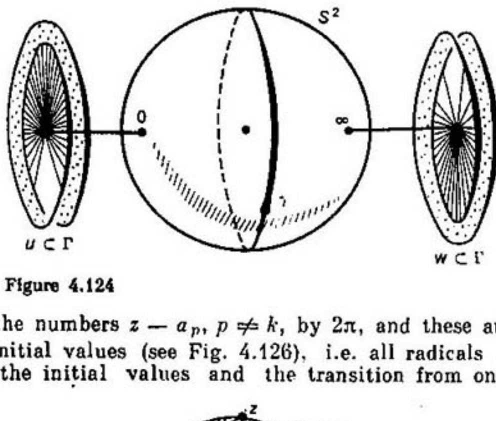

Figure 4.124

ments of the numbers $z-a_p$, $p\neq k$, by 2π, and these arguments take the initial values (see Fig. 4.126), i.e. all radicals $\sqrt{z-a_p}$ also take the initial values and the transition from one branch

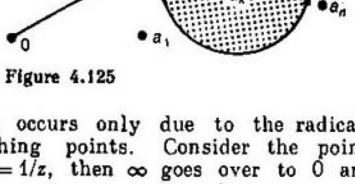

Figure 4.125

to another occurs only due to the radical $\sqrt{z-a_k}$. Thus, $\{a_k\}$ are branching points. Consider the point ∞ and make the change $u=1/z$, then ∞ goes over to 0 and $w=g(z)$ takes the

form $\frac{1}{\sqrt{u^n}}\prod_{k=1}^{n}\sqrt{1-a_k u}$. If n is even, circumvention round the point 0 returns the function to the initial value, so that 0 is not a branching point. If $n = 2p+1$ is odd, $w = \frac{1}{u^p}\frac{1}{\sqrt{u}}\times \prod_{k=1}^{n}\sqrt{1-a_k u}$ and 0 is a branching point. The proposition is proved.

As we arleady know, the surface Γ is the domain of single-valuedness of an algebraic function. Represent Γ as the gluing of several

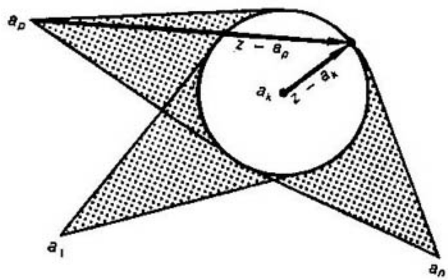

Figure 4.126

"sheets," each being the graph of a single-valued algebraic function. We start with an example.

Let us consider the function $w = \pm\sqrt{z}$ which is defined on S^2 and has two branching points, 0 and ∞. Connect the points 0 and ∞ by a smooth curve γ without self-intersections and make a cut along γ from 0 to ∞. Then two subsets, Γ_1 and Γ_2, arise on S^2: Γ_1 is the graph of the branch $\varphi_1 = \sqrt{z}$ defined on $S^2\setminus\gamma$ and Γ_2 is the graph of the branch $\varphi_2 = -\sqrt{z}$ defined on $S^2\setminus\gamma$. Since γ connects the branching points (and there are no other branching points), circumventions round these points are prohibited, i.e. we cannot pass from φ_1 to φ_2, remaining on $S^2\setminus\gamma$. This implies that each of the sheets Γ_1 and Γ_2 is homeomorphic to $S^2\setminus\gamma$, and this homeomorphism is realized by the functions φ_1 and φ_2 (Γ_i is the graph of φ_i) (see Fig. 4.127). The projection π: $\mathbf{C}^2 \to \mathbf{C}^1(z)$ along $\mathbf{C}^1(w)$ maps homeomorphically each of the sheets Γ_1 and Γ_2 onto $S_1\setminus\gamma$. Thus, Γ is glued of two pieces, namely, the sheets Γ_1 and Γ_2. How can we realize

this gluing? Define on S^2 an orientation and mark one border with "+" and the other with "−". These marks appear on the borders of both sheets, Γ_1 and Γ_2. Since these sheets are homeomorphic to $S^2 \setminus \gamma$ reconstruction of Γ requires that two copies of $S^2 \setminus \gamma$ should be glued with an allowance for the position of the marks "+" and "−". Since the branches interchange in the circumvention round the branching points, the border γ_1^+ of the sheet Γ_1 should be glued to the

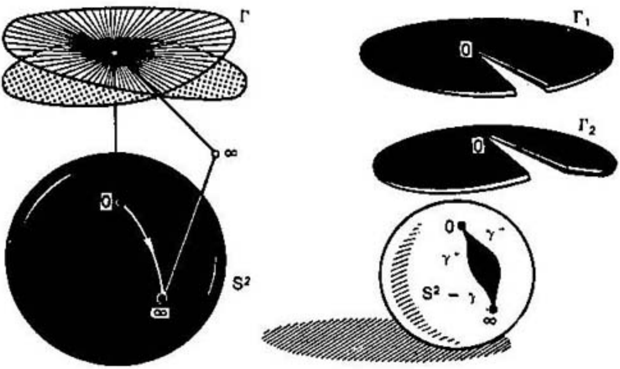

Figure 4.127

border γ_2^- of the sheet Γ_2; the borders γ_1^- and γ_2^+ of Γ_1 and Γ_2 are glued similarly (see Fig. 4.128). Apparently, such a gluing produces a sphere. Thus, we have proved the following statement.

Lemma 2. *The compactified Riemannian surface $\widetilde{\Gamma}$ of an algebraic function $w = \pm\sqrt{z}$ is homeomorphic to a sphere S^2.*

Let us consider the Riemannian surface of the function $w = \sqrt{(z-a)\cdot(z-b)}$, where $a \neq b$. According to Proposition 2, this function has two branching points: $z = a$ and $z = b$; though ∞ is not a branching point, it is, nevertheless, a singular point of the compactified surface $\widetilde{\Gamma}$, for at this point the two sheets touch each other and the values of both branches, φ_1 and φ_2, at infinity coincide (and are equal to infinity). Thus, ∞ is a singular point of the immersion of $\widetilde{\Gamma}$ in $\mathbf{C}^2 \cup (S^2 \vee S^2) = S^2 \times S^2$, but it is not a singular point of the surface $\widetilde{\Gamma}$ itself. Let us study the surface $\widetilde{\Gamma}$, assuming that the two sheets do not touch at ∞, i.e. we as if "unglue" the point ∞, thereby doubling it and providing each sheet with its "own" point ∞.

Lemma 3. *The compactified Riemannian surface $\widetilde{\Gamma}$ (with "unglued infinity") of the algebraic function $w = \pm\sqrt{(z-a)(z-b)}$ is homeomorphic to a two-dimensional sphere.*

Proof. The branching points are a and b, so that one has to make a cut γ from a to b and repeat the reasoning of Lemma 2. The lemma is proved.

As was already noted, the immersion $S^2 = \widetilde{\Gamma} \to S^2 \times S^2$ glues two points, and therefore the true location of $\widetilde{\Gamma}$ in $S^2 \times S^2$ is as in

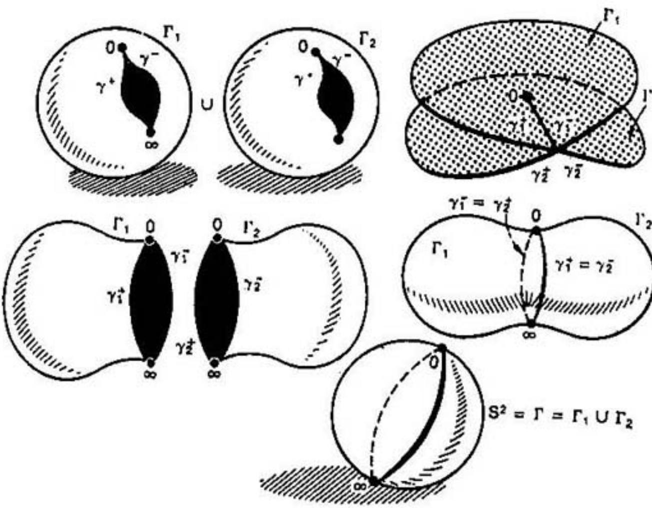

Figure 4.128

Fig. 4.129. Since we have proved that at the point ∞ the surface $\widetilde{\Gamma}$ is a smooth manifold, the singular point ∞ is the tangency of the two smooth sheets, rather than a conical point (see Fig. 4.130). Let us consider the general case.

Theorem 3. *Let $f(z, w) = w^2 - P_n(z)$, where the polynomial $P_n(z)$ does not have multiple roots. Then the compactified Riemannian surface $\widetilde{\Gamma}$ ("with unglued infinity") of the algebraic function $w = \pm\sqrt{P_n(z)} = \sqrt{\prod_{k=1}^{n}(z-a_k)}$ is homeomorphic to a sphere*

with $\left[\frac{n-1}{2}\right]$ *handles* ([ ] *denotes the whole part*), *i.e. to a manifold of type* M_g^2 *of genus* $g=\left[\frac{n-1}{2}\right]$. *The immersion of this sur-*

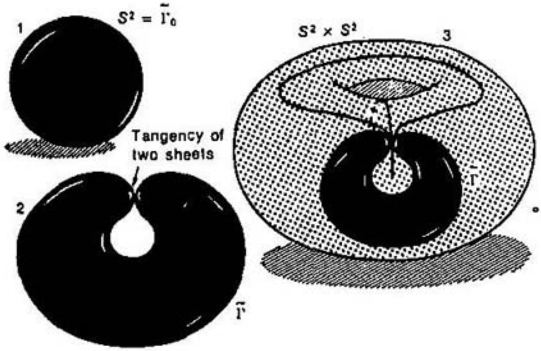

Figure 4.129

face in the compactified space $\mathbf{C}^2 \cup (S^2 \vee S^2) = S^2 \times S^2$ *is an embedding if the degree* n *is odd and an immersion gluing two points* ∞ *(on distinct sheets* Γ_1 *and* Γ_2*) if the degree* n *is even.*

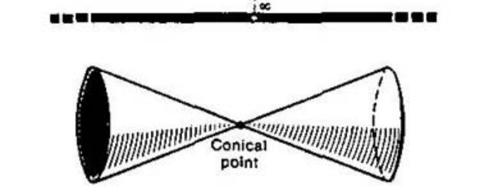

Figure 4.130

Proof. Let $n = 2p + 1$; according to Proposition 2, all the roots $a_1, \ldots, a_n$ and the point ∞ are branching points of $w=$

$\pm\sqrt{P_n(z)}=\sqrt{\prod_{h=1}^{p}(z-a_h)}$; this set can be subdivided into pairs, say (a_1, a_2), $\ldots$, (a_n, a_{n+1}), where $a_{n+1}=\infty$. Connecting the points a_{2k-1} and a_{2k} in each pair by a smooth segment γ_k, $1\leqslant k\leqslant p+1$, we obtain a sphere S^2 with $p+1$ cuts $\gamma_1, \ldots, \gamma_{p+1}$. Then each of the branches, φ_1 and φ_2, is a single-valued function on $S^2\setminus(\gamma_1\cup\ldots\cup\gamma_{p+1})$. Indeed, we have prohibited circumventions round an odd number of branching points; with our scheme of cuts, we can only

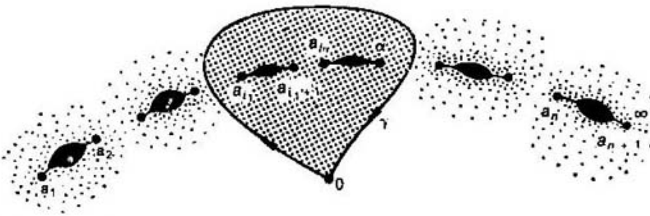

Figure 4.131

go round an even number of branching points. Suppose we have gone round the pairs $(a_{i_1}, a_{i_1+1}, \ldots, (a_{i_\alpha}, a_{i_\alpha+1})$. What does it happen to the branches φ_1 and φ_2? Since the argument increases by 2π only for those points that have been circumvented, the signs of the radicals change an even number of times. Because we deal with an even number of points, we need not necessarily distinguish between the concepts "inside the contour" and "outside the contour", for the number of singularities outside the contour is also even (furthermore, on a sphere the concepts "outside" and "inside" are dual) (see Fig. 4.131). In the circumvention along γ we have

$$\varphi=\prod_{t=1}^{\alpha}\sqrt{(z-a_{i_t})(z-a_{i_{t+1}})}\cdot\prod_{h\neq(i_1,\ldots,i_\alpha+1)}\sqrt{z-a_h}$$

$$\rightarrow\prod_{t=1}^{\alpha}\left(-\sqrt{z-a_{i_t}}\right)\left(-\sqrt{z-a_{i_{t+1}}}\right)\cdot\prod_{h\neq(i_1,\ldots,i_\alpha+1)}\sqrt{z-a_h}=\varphi.$$

If we choose another way of dividing the branching points into pairs, the domain obtained by cutting along the new paths will be homeomorphic to the domain considered above. Thus, the procedure does not depend on the method of grouping the branching points into pairs and on the way of connecting points in each pair by a smooth curve. Mark the borders of each cut with " + " and " − "; then each of the sheets, Γ_1 and Γ_2, is homeomorphic to $S^2\setminus(\gamma_1\cup\ldots\cup\gamma_{p+1})$.

Hence, to reconstruct Γ, one has to glue these two sheets together by attaching the border $\gamma^+_{i(1)}$ on Γ_1 to the border $\gamma^-_{i(2)}$ on Γ_2 (Fig. 4.132). We obtain M^2_g, i.e. sphere S_1 with $g = p = \left[\frac{n-1}{2}\right]$ handles.

Let $n = 2p$. According to Proposition 2, only the roots $a_1, \ldots, a_{2p}$ are branching points; at infinity there exists a singular point,

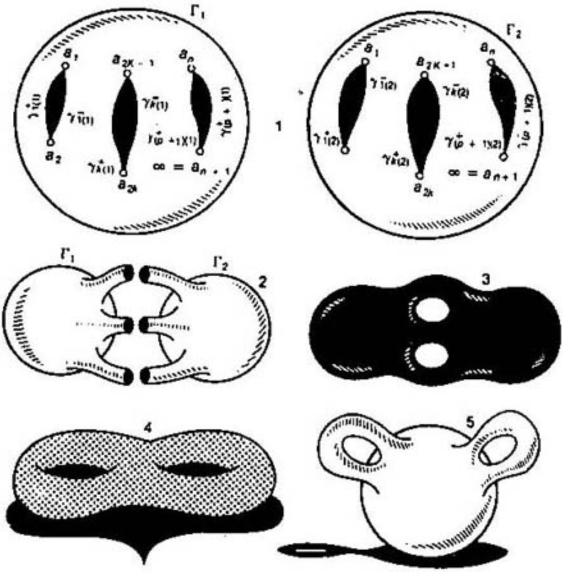

Figure 4.132

namely, the tangency of two sheets when $\widetilde{\Gamma}$ is immersed in $S^2 \times S^2$. Again, dividing the roots into pairs, connecting the points in each pair by smooth segments, marking the borders of the cuts, and cutting $\widetilde{\Gamma}$ along the inverse images of these segments, we obtain two sheets. The point ∞ is "unglued" and each sheet is provided with its own point ∞. Finally, $\widetilde{\Gamma}$ is reconstructed by making gluing in the inverse order (see Fig. 4.133). The theorem is proved.

Thus, two-dimensional surfaces arise quite naturally: as Riemannian surfaces of certain algebraic functions.

Statement 1. *Each two-dimensional smooth, compact, connected, closed manifold of genus g (i.e. of type M_g^2) can be endowed with the structure of a complex-analytic manifold.*

Proof. Let us consider Γ for the function $w = \pm\sqrt{P_n(z)}$, where P_n has simple roots. Since $\Gamma \subset \mathbf{C}^2$, where it is defined by the graph $w = g(z)$ or by $z = \omega(w)$, we have two orthogonal projections

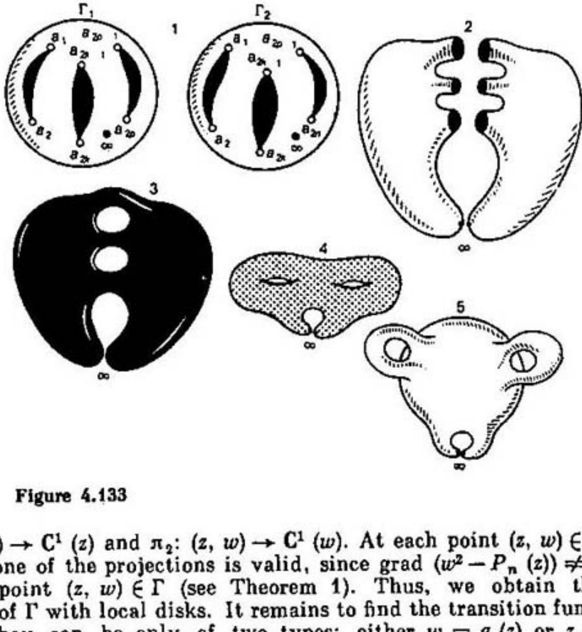

Figure 4.133

π_1: $(z, w) \to \mathbf{C}^1(z)$ and π_2: $(z, w) \to \mathbf{C}^1(w)$. At each point $(z, w) \in \Gamma$ at least one of the projections is valid, since grad $(w^2 - P_n(z)) \neq 0$ at each point $(z, w) \in \Gamma$ (see Theorem 1). Thus, we obtain the covering of Γ with local disks. It remains to find the transition functions. They can be only of two types: either $w = g(z)$ or $z = \omega(w)$. Since the function $g(z)$ or $\omega(w)$ is complex-analytic (see Proposition 1), the statement is proved, because the continuation of the complex structure to the point ∞ of the surface $\tilde{\Gamma}$ holds inside the complex-analytic manifold $S^2 \times S^2 = \mathbf{CP}^1 \times \mathbf{CP}^1$.

It is a simple matter to demonstrate that a manifold M_μ^2 cannot be endowed with the structure of a complex-analytic manifold (verify!).

Statement 2. *Each two-dimensional smooth, compact, connected, closed manifol of genus g (i.e. of type M_g^2) can be provided with a conformal Riemannian metric.*

Proof. The metric is called conformal if there exist local coordinates in which this metric is of the form $g_{ij}=\alpha\cdot\delta_{ij}$. Let us realize M_g^2 in (extended) $\mathbf{C}^2$ as the Riemannian surface of the function

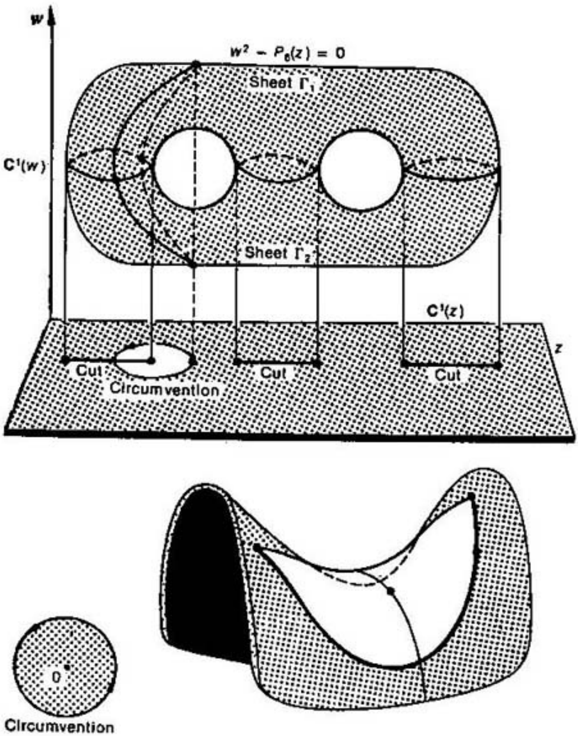

Figure 4.134

$w = \pm\sqrt{P_n(z)}$. Consider on M_g^2 the coordinate z which runs $\mathbf{C}^1(z)$. This complex coordinate is valid in the entire M_g^2 except at the point 0 near which we should introduce the coordinate w related to z by the complex-analytic transformation $w = g(z)$. Introduce in

$C^2 = R^4$ the metric

$$ds^2 = dz\,d\bar{z} + dw\,d\bar{w} = \sum_{k=1}^{4} (dx^k)^2$$

and consider its restriction to $\Gamma \subset C^2$. We have

$$ds^2(M_g^2) = dzd\bar{z} + dg(z)\,\overline{dg(z)}$$
$$= dzd\bar{z} + g'(z)\,\bar{g}'(z)\,dzd\bar{z} = (1 + |g_z'(z)|^2)\,dzd\bar{z}$$

or $ds^2(M_g^2) = (1 + u_x^2 + v_x^2)(dx^2 + dy^2)$, where $z = x + iy$, $g = u + iv$. Here

$$g'(z) = \frac{d}{dz} g(z) = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)(u + iv)$$
$$= \frac{1}{2}(u_x + iv_x - iu_y + v_y) = u_x + iv_x = v_y + iv_x,$$

because $u_x = v_y$ and $u_y = -v_x$. The statement is proved.

If we make the complex-analytic change $z = \eta(\xi)$, the conformal metric derived above remains conformal, since

$$ds^2 + \alpha(z)\,dzd\bar{z} = \alpha(\eta(\xi))\,|\eta'(\xi)|^2\,d\xi d\bar{\xi} = \rho(\xi)\,d\xi d\bar{\xi}.$$

In conclusion, a few words about the way Γ is embedded in C^2 in a neighbourhood of a branching point. Though we have proved that at such a point Γ is a smooth manifold, it appears difficult to depict a branching point in R^3, because the space R^3 has "low dimension" as compared to $R^4 = C^2$. Figure 4.134 is an attempt to illustrate schematically circumvention round a branching point.

Chapter 5

Tensor Analysis and Riemannian Geometry

5.1. THE GENERAL CONCEPT OF A TENSOR FIELD ON A MANIFOLD

Let us consider a smooth manifold M^n. Let $P \in M^n$ and let $x^1, \ldots, x^n$ be a local regular coordinate system in a neighbourhood of P. There also exist infinitely many other regular coordinate systems in the neighbourhood of P. Our task is to study the objects which are invariant under different coordinate transformations. The new coordinates will be denoted by a prime: $x^{1'}, \ldots, x^{n'}$.

(1) Let $\mathbf{a} \in T_P(M^n)$ be a tangent vector to M^n at a point P. Previously, we have given several definitions of a tangent vector; some of them defined this vector irrespective of the choice of local coordinates, i.e. as an object invariant under regular coordinate transformations (viz. definition via the sheaf of tangent curves). However, if we want to associate with a tangent vector its coordinates, i.e. to express this vector analytically, we need to introduce a coordinate system. If we choose another system, the coordinates will change. Let us find relations between the coordinates of a vector in different systems. Since

$\mathbf{a} = a^i \frac{\partial}{\partial x^i} = a^{i'} \frac{\partial}{\partial x^{i'}}$, we have $a^{i'} \frac{\partial}{\partial x^{i'}} = a^i \frac{\partial x^{i'}}{\partial x^i} \frac{\partial}{\partial x^{i'}}$ and since $\left\{\frac{\partial}{\partial x^{i'}}\right\}$ form a basis in $T_P(M^n)$, $a^{i'} = \frac{\partial x^{i'}}{\partial x^i} a^i$. Thus, we have found how the coordinates of a vector change under system transformation: the coordinates $a^1, \ldots, a^n$ are transformed by the Jacobi matrix $\left(\frac{\partial x^{i'}}{\partial x^i}\right) = J$. We have obtained this condition from the requirement that a vector, as a geometric object, be invariant under coordinate transformation.

On the contrary, we can define a vector as an object given in each coordinate system $x^1, \ldots, x^n$ by a set of numbers $a^1, \ldots, a^n$ which are transformed while passing from the system $x^1, \ldots, x^n$ to another

system $x^{1'}, \ldots, x^{n'}$ according to the law $a^{i'} = \frac{\partial x^{i'}}{\partial x^i} a^i$. Apparently, the object thus defined is invariant.

(2) Let $f(x)$ be a smooth function on M^n; consider grad $f = \left\{\frac{\partial f}{\partial x^i}\right\}$. This set of numbers (functions of a point P) is valid in each coordinate system. Let us find the transformation law. According to the rule of differentiation of a composite function,

$$\xi_{i'} = \frac{\partial f}{\partial x^{i'}} = \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial x^{i'}} = \xi_i \frac{\partial x^i}{\partial x^{i'}},$$

where

$$\text{grad } f = \{\xi_1, \ldots, \xi_n\}, \quad \xi_i = \frac{\partial f}{\partial x^i}, \quad \left(\frac{\partial x^i}{\partial x^{i'}}\right) = (J^T)^{-1},$$

i.e. the set $(\xi_1, \ldots, \xi_n)$ is transformed in a different way than the set of the components of a vector: namely, grad f is transformed by the matrix $(J^{-1})^T$, where J is the Jacobi matrix $\frac{\partial x^{i'}}{\partial x^i}$. This law can also be deduced through the requirement that a geometric object, the derivative of a function f with respect to a vector $\mathbf{a} \in T_P(M^n)$, be invariant under coordinate transformation. Indeed, this derivative does not depend on the choice of local coordinates, so that

$$\frac{df}{d\mathbf{a}} = \lim_{\varepsilon \to 0} \frac{f(\gamma(\varepsilon)) - f(\gamma(0))}{\varepsilon},$$

where

$$\gamma(t) \in M^n, \quad \gamma(0) = P, \quad \dot{\gamma}(0) = \mathbf{a} \in T_p(M^n),$$
$$\frac{df}{d\mathbf{a}} = \frac{\partial f}{\partial x^i} \frac{dx^i(t)}{dt}\bigg|_{t=0} = a^j \frac{\partial f}{\partial x^i} = a^{i'} \frac{\partial f}{\partial x^{i'}}.$$

Thus,

$$a^{i'} \frac{\partial f}{\partial x^{i'}} = a^i \frac{\partial x^{i'}}{\partial x^i} \frac{\partial f}{\partial x^{i'}}, \text{ i.e.}$$

$$a^{i'}\xi_{i'} = a^i \frac{\partial x^{i'}}{\partial x^i} \xi_{i'} = a^i \xi_i, \quad \xi_i = \xi_{i'} \frac{\partial x^{i'}}{\partial x^i}, \quad \xi_{i'} = \frac{\partial x^i}{\partial x^{i'}} \xi_i,$$

which is what was required.

(3) Let us consider $T_P(M^n)$ and the adjoint space $T_P^*(M^n)$ of real-valued linear functionals $l(\mathbf{a})$, where $\mathbf{a} \in T_P(M^n)$. We now find the transformation law for the coordinates of these functionals under coordinate substitution in some neighbourhood of a point; introduce in $T_P^*(M^n)$ a basis $\mathbf{e}^1, \ldots, \mathbf{e}^n$ consisting of the functionals satisfying the condition $e^k(\mathbf{e}_\alpha) = \delta_\alpha^k$, where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis in $T_P(M^n)$.

Since l (a) is a real number and since the scalar does not depend on the choice of coordinates, $l\,(\mathbf{a})_{(x)} = l'\,(\mathbf{a}')_{(x')}$, whence

$$\mathbf{a}=a^i\mathbf{e}_i,\ l\,(\mathbf{a})=(l_ie^i)\,(a^k\mathbf{e}_k)=l_ka^k=l_{k'}a^{k'}=l_{k'}\,\frac{\partial x^{k'}}{\partial x^k}\,a^k.$$

Since $\mathbf{a}$ is arbitrary, we have from these equalities $l_k=\frac{\partial x^{k'}}{\partial x^k}\,l_{k'}$, i.e. $l_{k'}=\frac{\partial x^k}{\partial x^{k'}}\,l_k$, hence, the coordinates of the functional are transformed exactly in the same way as the coordinates of grad f. The elements of $T^*_P\,(M^n)$ are called covectors; thus, grad f is a covector.

We have introduced in T^* the adjoint basis consisting of covectors $\mathbf{e}^1, \ldots, \mathbf{e}^n$, so that T can be identified with T^* by setting $\varphi\colon T \to T^*$; $\varphi\,(\mathbf{a}) = l$, where $\mathbf{a} = a^k\mathbf{e}_k$; $l = \sum\limits_{k=1}^{n} a^k e^k$. Hence, we assume that l mapped by the isomorphism φ into $\mathbf{a}$ has, relative to $\mathbf{e}^1, \ldots, \mathbf{e}^n$, the same coordinates as $\mathbf{a}$ relative to $\mathbf{e}_1, \ldots, \mathbf{e}_n$. This linear isomorphism is not invariant under the coordinate substitution $x \to (x')$, since a vector and a covector obey different transformation laws: a vector is transformed by the matrix J and a covector by the matrix $(J^{-1})^T$. Thus, the identification of T with T^* valid in the system (x) becomes invalid under an arbitrary coordinate substitution $(x) \to (x')$. There are, however, transformations which preserve the correspondence $\varphi\colon T \to T^*$, for these transformations $J = (J^{-1})^T$, i.e. they are defined at a given point P by an orthogonal Jacobi matrix. A general transformation has a non-singular Jacobi matrix, rather than an orthogonal one. Nevertheless, T and T^* can be identified by a linear isomorphism which works for any coordinate substitution. This topic is discussed below.

(4) Let C be a linear homogeneous operator which turns $T_P\,M^n$ into itself and let the same letter stand for the matrix of the operator written in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Making the substitution $(x) \to (x')$, we obtain $C' = JCJ^{-1}$, i.e. $c^{i'}_{j'} = \frac{\partial x^{i'}}{\partial x^i}\,\frac{\partial x^j}{\partial x^{j'}}\,c^i_j$. The transformation law for the operator matrix can be deduced from the requirement that the relation $\mathbf{b} = C\,(\mathbf{a})$ ($\mathbf{a}$ is an arbitrary vector in T_PM^n) be invariant under coordinate transformation. Indeed, all the objects in this relation admit invariant (i.e. independent of the coordinates) definition which, apparently, leads to the desirable transformation law for C. The proof of this statement is left to the reader as an exercise.

(5) Let us consider a bilinear symmetric form B on $T_P\,(M^n)$; $B\,(\mathbf{a}, \mathbf{b}) = b_{ij}a^ib^j$. It is known from algebra that under a coordinate substitution the matrix B is transformed into a matrix B' such that

$B' = JbJ^T$, i.e. $b_{i'j'} = \frac{\partial x^i}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} b_{ij}$. As before, this transformation law for B can be obtained from the requirement that the scalar $B(\mathbf{a}, \mathbf{b})$, $\mathbf{a}, \mathbf{b} \in T_P$ be invariant under an arbitrary coordinate substitution. Indeed,

$$B(\mathbf{a}, \mathbf{b}) = b_{ij}a^ib^j = b_{i'j'}a^{i'}b^{j'} = b_{ij}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^{j'}}{\partial x^j} a^ib^j \text{ i.e.}$$

$$b_{ij} = b_{i'j'}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^{j'}}{\partial x^j}, \text{ or } b_{i'j'} = \frac{\partial x^i}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}} b_{ij}.$$

We have considered simple examples of transformation laws; their distinction was indicated by different indexation of the object components: a^i, l_j, c^i_j, b_{ij}. The superscripts are called *contravariant* indices and the subscripts *covariant* indices. It is convenient to associate with each object a pair (p, q), where q is the number of covariant indices and p is the number of contravariant indices. Thus, the set of coordinates for a vector $\mathbf{a} \in T_PM^n$ is of the type $(1, 0)$, for a covector of the type $(0, 1)$, for an operator of the type $(1, 1)$ and for a bilinear form on vectors of the type $(0, 2)$. Besides this form, there exists the form $B(l, m)$ on the covectors $l, m \in T_P^*M^n$. Calculation (by analogy with $B(\mathbf{a}, \mathbf{b})$ shows that if $B(l, m) = b^{ij}l_im_j$, then $b^{i'j'} = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^{j'}}{\partial x^j} b^{ij}$ (verify!), i.e. $B' = (J^{-1})^TB(J^{-1})$. The objects just mentioned admit an invariant, coordinate-independent definition in a particular coordinate system, and the transformation law for the components of these objects directly follows from the invariance of this definition. Thus, if in each coordinate system there is given a set of numbers (functions) transformed (under coordinate substitution) according to the above laws, we may say that these sets define an invariant object which can be studied through its coordinates in different coordinate systems. This circumstance underlies the general definition of a tensor field on a manifold as an invariant geometric object which does not depend on the choice of the local coordinate system. A particular expression of the components of this object does depend on the coordinate system.

Definition. *By a tensor of type (p, q) of rank $p + q$ (correspondingly, a tensor field of type (p, q) of rank $p + q$) we mean an object defined in each coordinate system $(x) = (x^1, \ldots, x^n)$ by the set of numbers $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ (correspondingly, by the set of smooth functions $T^{i_1\ldots i_p}_{j_1\ldots j_q}(x)$) transformed under the coordinate substitution $(x) \to (x')$ according to the law*

$$T^{i'_1\ldots i'_p}_{j'_1\ldots j'_q} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_p}}{\partial x^{i_p}} \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_q}}{\partial x^{j'_q}} T^{i_1\ldots i_p}_{j_1\ldots j_q}.$$

This definition satisfies the requirement that an object can be defined invariantly, i.e. irrespective of its expression in a particular coordinate system. Indeed, it is known from algebra that a tensor (tensor field) can be defined as a polylinear mapping

$$T : \underbrace{T_P \times \ldots \times T_P}_{q} \times \underbrace{T_P^* \times \ldots \times T_P^*}_{p} \to \mathbf{R},$$

defined by the relation

$$T(\mathbf{a}_1, \ldots, \mathbf{a}_q, l_1, \ldots, l_p) = T^{i_1 \ldots i_p}_{j_1 \ldots j_q} a_1^{j_1} \ldots a_q^{j_q} l^1_{i_1} \ldots l^p_{i_p},$$

where $\mathbf{a}_k = a_k^{j_k} \mathbf{e}_{j_k}$, $l^k = l^k_{i_k} e^{i_k}$, $\{a_k^{j_k}\}$ are the coordinates of the vector $\mathbf{a}_k \in T_P(M^n)$ (the kth factor), $\{l^k_{i_k}\}$ are the coordinates of the covector $l^k \in T_P^*(M^n)$ (the kth factor). The functions $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ are the coefficients of the mapping T. It is known from algebra that a multilinear mapping can always be written in this form. If the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $e^1, \ldots, e^n$ are fixed, we have a useful formula for $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$

$$T^{i_1 \ldots i_p}_{j_1 \ldots j_q} = T(e_{j_1}, \ldots, e_{j_q}, e^{i_1}, \ldots, e^{i_p}).$$

Indeed,

$$\begin{aligned} T(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_q}, e^{i_1}, \ldots, e^{i_p}) &= T^{\alpha_1 \ldots \alpha_p}_{\rho_1 \ldots \rho_q} \cdot (\mathbf{e}_{j_1})^{\rho_1} \cdot \ldots \cdot (\mathbf{e}_j)^{\rho_q} (e^{i_1})_{\alpha_1} \cdot \ldots \cdot (e^{i_p})_{\alpha_p} \\ &= T^{\alpha_1 \ldots \alpha_p}_{\rho_1 \ldots \rho_q} \delta^{\rho_1}_{j_1} \cdot \ldots \cdot \delta^{\rho_q}_{j_q} \cdot \delta^{j_1}_{\alpha_1} \cdot \ldots \cdot \delta^{i_p}_{\alpha_p} = T^{i_1 \ldots i_p}_{j_1 \ldots j_q}. \end{aligned}$$

Furthermore, T can be defined in invariant terms, since all the objects appearing in the formula $T: (T_P)^q \times (T_P^*)^p \to \mathbf{R}$, as well as the multilinearity property, are invariant. A particular expression for the coefficients of $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ depends on the choice of the coordinates.

Although the invariant form of a multilinear mapping is more convenient, we shall almost always consider the "coordinate form" because it appears quite useful in particular calculations. Let us introduce, for convenience, a special notation, the so-called multi-indices: $(i) = (i_1, \ldots, i_p)$ where the entire set of indices is replaced by a single one. Then, the tensor law is written as

$$T^{(i')}_{(j')} = \frac{\partial x^{(i')}}{\partial x^{(i)}} \frac{\partial x^{(j)}}{\partial x^{(j')}} T^{(i)}_{(j)}, \text{ where } \frac{\partial x^{(i')}}{\partial x^{(i)}} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_p}}{\partial x^{i_p}}.$$

We now turn to studying simple properties of tensor fields.

Lemma 1. *Let* $(x) \to (x')$ *be a regular substitution. Then the relation* $T^{(i')}_{(j')} = \frac{\partial x^{(i')}}{\partial x^{(i)}} \frac{\partial x^{(j)}}{\partial x^{(j')}} T^{(i)}_{(j)}$ *(the tensor law) implies that*

$$T^{(i)}_{(j)} = \frac{\partial x^{(i)}}{\partial x^{(i')}} \frac{\partial x^{(j)}}{\partial x^{(j')}} T^{(i')}_{(j')}.$$

The *proof* of the lemma follows from the identity $\frac{\partial x^i}{\partial x^{h'}} \frac{\partial x^h}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i_j$ which means that $J \cdot J^{-1} = E$, where $J = \left(\frac{\partial x^{i'}}{\partial x^i}\right)$ is the Jacobi matrix.

Lemma 2. *Given a tensor field written in the system* (x), *and let the coordinate substitutions* $(x) \to (z) \to (y)$ *and* $(x) \to (v) \to (y)$ *be valid, i.e. we pass from* (x) *to* (y) *in two ways: via* (z) *and via* (v). *Then the expression for the field* T *in* (y) *does not depend on the way* (x) *is transformed into* (y).

The *proof* follows from the formula for differentiation of a composite function

$$\frac{\partial y^i}{\partial x^k} = \frac{\partial y^i}{\partial z^q} \frac{\partial z^q}{\partial x^k} = \frac{\partial y^i}{\partial v^\alpha} \frac{\partial v^\alpha}{\partial x^k}.$$

Let us recall some properties of tensors known from aglebra.

(a) Tensors of rank $p + q$ and type (p, q) (they are considered at a point) form a linear space H such that $\dim H = n_{p+q}$. The proof follows from the representation of a tensor as a multilinear mapping: the linear combination of such mappings is also multilinear.

(b) In the space H of all tensors of rank $p + q$ and type (p, q) there exists an additive basis; the elements of this basis (multilinear mappings) are denoted by

$$\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}.$$

The number of these elements is equal to n^{n+p} (verify!). Each mapping

$$\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}$$

is given by the formulas

$$\mathbf{e}_{i_\alpha}(l) = l_{i_\alpha}, \text{ where } l \in T^*_p(M^n), \quad e^{j_\alpha}(\mathbf{a}) = a^{j_\alpha}, \quad \mathbf{a} \in T_pM^n,$$

$$\begin{aligned}(\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}) & (\mathbf{a}_1, \ldots, \mathbf{a}_q l^1, \ldots, l^p) \\ &= \mathbf{e}_{i_1}(l^1) \cdot \ldots \cdot \mathbf{e}_{i_p}(l^p) \cdot e^{j_1}(\mathbf{a}_1) \cdot \ldots \cdot e^{j_q}(\mathbf{a}_q) \\ &= l^1_{i_1} \cdot \ldots \cdot l^p_{i_p} \cdot a^{j_1}_1 \cdot \ldots \cdot a^{j_q}_q).\end{aligned}$$

If $T\colon (T_P M^n)^q \times (T^*_P M^n)^p \to \mathbf{R}$ is an arbitrary multilinear map-

ping of type (p, q), then

$$T(\mathbf{a}_1, \ldots, \mathbf{a}_q, l^1, \ldots, l^p) = T^{i_1\ldots i_p}_{j_1\ldots j_q} a_1^{j_1} \ldots a_q^{j_q} l^1_{i_1} \ldots l^p_{i_p}$$

$$= T^{i_1\ldots i_p}_{j_1\ldots j_q} (\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}) (\mathbf{a}_1, \ldots, \mathbf{a}_q, l^1, \ldots, l^p),$$

i.e.

$$T = T^{i_1\ldots i_p}_{j_1\ldots j_q} (\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}).$$

Thus, the mappings $(\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \mathbf{e}^{j_1} \otimes \ldots \otimes \mathbf{e}^{j_q})$ form an additive basis in the space of all tensors of type (p, q). With the aid of multi-indices, this decomposition of an arbitrary tensor T in the basis tensors can be written as $T = T^{(i)}_{(j)} \mathbf{e}_{(i)} \otimes e^{(j)}$. The transformation law for the basis tensors $\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \ldots \otimes e^{j_q}$ under coordinate substitution is of the form

$$\mathbf{e}_{(i')} \otimes e^{(j')} = \frac{\partial x^{(i)}}{\partial x^{(i')}} \frac{\partial x^{(j')}}{\partial x^{(j)}} \mathbf{e}_{(i)} \otimes e^{(j)}$$

(verify!). All these properties and definitions are transferred to the case of smooth tensor fields on a manifold; the difference is that here we should consider tensor fields as linear combinations of basis fields of type (p, q) with variable coefficients, so that the space of tensor fields becomes an infinite-dimensional space.

A particular case: any tensor field of type (1, 0) admits (locally) the representation $\mathbf{T} = T^i(x)\, \mathbf{e}_i(x)$; similarly, any field of type (0, 1) admits the representation $l = l_i e^i(x)$.

5.2. SIMPLE TENSOR FIELDS

5.2.1. EXAMPLES

(1) Let a vector field $\mathbf{T}(x)$ be defined on M^n. Then $\mathbf{T}(x)$ is a tensor field of type (1, 0); the space of all smooth fields of type (1, 0) is identified with the infinite-dimensional space of all smooth vector fields on M^n.

(2) Let a smooth covector field be defined on M^n, i.e. at each point $x \in M^n$ there is fixed a linear functional $l(x) \in T^*_x M^n$ smoothly dependent on the point. This field is an element of the infinite-dimensional space of tensor fields of type (0, 1). This space contains a linear infinite-dimensional subspace of the fields grad $f(x)$, where f is a smooth function on M^n.

(3) Let the Riemannian metric $g_{ij}(y)$ be given on M^n. Since $g_{i'j'} = \frac{\partial x^i}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} g_{ij}$, the set of all $g_{ij}(x)$ defines a tensor field of type (0, 2), i.e. a smooth field of bilinear forms. This tensor is called a *metric tensor*.

(4) A complex-analytic manifold M^{2n} is characterized by the property that in each T_xM^{2n} there is defined the operator I of "multiplication by imaginary unity", $I^2 = -E$. Since this operator smoothly depends on the point, we obtain a tensor field of type (1, 1).

(5) *Inertia tensor*. Let us consider in $\mathbf{R}^3$ arigid body with one fixed point 0, the body can rotate about the point 0. Suppose the body consists of a finite number N of material points rigidly connected with one another (their mutual position does not change during the motion). We shall only deal with orthogonal coordinate systems in $\mathbf{R}^3$. Let $m_1, m_2, \ldots, m_N$ stand for the masses of the points and let $x^i_{(1)}, \ldots, x^i_{(N)}$ $(i = 1, 2, 3)$ denote their coordinates at a fixed moment t. Let us consider the matrix

$$a^{ij} = -\sum_{\alpha=1}^{N} m_\alpha x^i_\alpha x^j_{(\alpha)} + \delta^{ij} \sum_{\alpha=1}^{N} m_\alpha \mid \mathbf{x}_{(\alpha)} \mid^2.$$

We obtain a tensor of rank 2 and type (0, 2), the matrix a^{ij} being symmetric. This symmetric tensor a^{ij} is called the inertia tensor of a rigid body (various bodies have distinct inertia tensors). The role of this tensor may be illustrated by the following example. Let us consider a straight line with the unit direction vector $\mathbf{l} = (l^1, l^2, l^3)$ through the point 0 and evaluate the sum $\Sigma_{ij} a^{ij} l^i l^j = H(\mathbf{l})$. Since a^{ij} and l^i are tensors, $H(\mathbf{l})$ is an invariant of orthogonal coordinate transformation, i.e. a scalar. Calculation yields

$$\begin{aligned} H(\mathbf{l}) &= \sum_{ij} a^{ij} l^i l^j \\ &= -\sum_{\alpha=1}^{N} m_\alpha \sum_i x^i_{(\alpha)} l^i \sum_j x^j_{(\alpha)} l^j + \sum_{ij} \delta^{ij} l^i l^j \sum_{\alpha=1}^{N} m_\alpha \mid \mathbf{x}_{(\alpha)} \mid^2 \\ &= -\sum_{\alpha=1}^{N} m_\alpha (\langle \mathbf{x}_{(\alpha)}, \mathbf{l} \rangle)^2 + \sum_{\alpha=1}^{N} \mid \mathbf{l} \mid^2 \mid \mathbf{x}_{(\alpha)} \mid^2 \\ &= \sum_{\alpha=1}^{N} m_\alpha (\mid \mathbf{x}_{(\alpha)} \mid^2 - (\langle \mathbf{x}_{(\alpha)}, \mathbf{l} \rangle)^2). \end{aligned}$$

This expression can be interpreted as follows: the factor $\mid \mathbf{x}_\alpha \mid^2 - \langle \mathbf{x}_\alpha, \mathbf{l} \rangle^2$ is the squared distance from the point of mass m_α to the axis $\mathbf{l}$. Thus, we have obtained the familiar expression for the moment of inertia of a body about an axis. The eigenvectors of the matrix a^{ij} are parallel to the so-called principal axes of inertia of a rigid body.

(6) *Strain tensor*. Let us consider a continuous medium which occupies a certain volume in $\mathbf{R}^n$ referred to Cartesian coordinates $x^1, \ldots$ $\ldots, x^n$. We identify the points of this volume with the points of

the medium and consider small deformation of the medium, viz., due to external forces. Suppose we are given displacements of the points, i.e. smooth functions $u^i(x^1, \ldots, x^n)$, $1 \leqslant i \leqslant n$, which depend on the point and enable the coordinates of the displaced point $\widetilde{P}$ to be calculated in terms of the coordinates of the initial point P by the formula $\widetilde{x}^i = x^i + u^i(x^1, \ldots, x^n)$. We shall consider small displacements, i.e. the functions $u^i(x)$ are assumed to be rather small. To formulate the problem correctly, one should deal

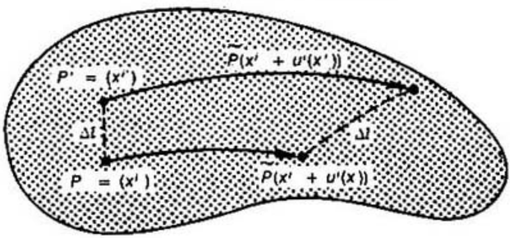

Figure 5.1

with infinitesimal displacements, which is equivalent to defining a smooth vector field in a domain, though the theory of continuous media sometimes handles finite deformations. We shall use the formalism suggested above. Thus, suppose a point $P = (x^1, \ldots, x^n)$ shifts to the point $\widetilde{P} = (x^i + u^i(x^1, \ldots, x^n))$. How does the length of a smooth curve change under medium deformation? Consider two nearby points $P(x^1, \ldots, x^n)$ and $P'(x^{1'}, \ldots, x^{n'})$. Let $(\Delta l)^2 = \sum_{i=1}^{n} (x^{i'} - x^i)^2 = \sum_{i=1}^{n} (\Delta x^i)^2$ be the length of the Euclidean segment connecting P and P'; find the length $(\Delta l')^2$ of the segment connecting the images $\widetilde{P}$ and $\widetilde{P}'$ of the points P and P' after their displacement as a result of medium deformation (Fig. 5.1). We have

$$\begin{aligned}
(\Delta l')^2 &= \sum_{i=1}^{n} (x^{i'} + u^i(x') - x^i - u^i(x))^2 \\
&= \sum_{i=1}^{n} [(x^{i'} - x^i) + (u^i(x') - u^i(x))]^2 = \sum_{i=1}^{n} (\Delta x^i + \Delta u^i)^2 \\
&= \sum_{i=1}^{n} (\Delta x^i)^2 + 2\sum_{i=1}^{n} \Delta x^i \Delta u^i + \sum_{i=1}^{n} (\Delta u^i)^2 \\
&= (\Delta l)^2 + 2\sum_{i=1}^{n} \Delta x^i \Delta u^i + \sum_{i=1}^{n} (\Delta u^i)^2.
\end{aligned}$$

Since $\Delta u^i = \frac{\partial u^i}{\partial x^k} \Delta x^k$, we have

$$(\Delta l')^2 - (\Delta l)^2 = 2 \sum_{i=1}^{n} \Delta x^i \Delta u^i + \sum_{i=1}^{n} (\Delta u^i)^2 = 2 \sum_{i,\,k} \frac{\partial u^i}{\partial x^k} \Delta x^i \Delta x^k$$
$$+ \sum_{i,\,k,\,p} \frac{\partial u^i}{\partial x^k} \frac{\partial u^i}{\partial x^p} \Delta x^k \Delta x^p$$
$$= \sum_{i<k} \left(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i} \right) \Delta x^i \Delta x^k$$
$$+ \sum_{i,\,k,\,p} \frac{\partial u^i}{\partial x^k} \frac{\partial u^i}{\partial x^p} \Delta x^k \Delta x^p.$$

Thus,

$$(dl')^2 - (dl)^2 = \sum_{i<k} \left(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i} \right) dx^i dx^k + \sum_{i,\,k,\,p} \frac{\partial u^i}{\partial x^k} \frac{\partial u^i}{\partial x^p} dx^k dx^p.$$

If the strains $u^i(x)$ are small, we can assume

$$(dl')^2 - (dl)^2 \cong \sum_{i<k} \left(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i} \right) dx^i dx^k.$$

Set $\eta_{ik}(x) = \frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i}$, then $(dl')^2 - (dl^2) \cong \eta_{ik} dx^i dx^k$. The set of functions η_{ik} forms a tensor of rank 2 of type (0, 2), called the strain tensor.

(7) *Stress tensor*. Let us consider an elastic body subject to a deformation which produces stresses in the body. To explain this term, we resort to the conventional model developed in the classical elasticity theory. Let $d\sigma$ be an elementary area through a point P in the body and let $\mathbf{n}(P)$ be the normal to $d\sigma$ at P. We shall assume that $d\sigma$ is oriented and the normal is also provided with an orientation (Fig. 5.2). In some neighbourhood of P the element $d\sigma$ divides the body into two parts: one is located on the positive side of $d\sigma$ and the other on the negative side. The existence of stresses in an (elastic) body suggests that there exists a force $\mathbf{F}$ exerted by the matter on one side of the element $d\sigma$ upon the matter of the other side; this force is applied at the point P. To describe stresses in a body usually means to find the force (due to a stress) exerted on an arbitrary oriented infinitesimal element (the shape of the element may be arbitrary). It is assumed that for a given normal $\mathbf{n}(P)$ the force $\mathbf{F}$ acting on $d\sigma$ is proportional to its area which is also denoted by $d\sigma$. Thus, we obtain a correspondence between $\mathbf{n}$ and $\mathbf{F}$: with each normal $\mathbf{n}(P)$ at P there is associated a vector $\mathbf{F}(\mathbf{n})$. The character of this dependence is established in classical elasticity theory pro-

ceeding from mechanical considerations. We shall not go into details, but only report the final result: it appears (in a good approximation to the real situation) that the dependence $\mathbf{F}(\mathbf{n})$ may be considered as linear, i.e. at each point P there is defined a linear operator $\mathbf{n} \to Q(\mathbf{n})$. Then $\mathbf{F}(\mathbf{n})$ can be written as $\mathbf{F}(\mathbf{n}) = Q_P(\mathbf{n})\, d\sigma$, where $F^i = Q^i_j n^j d\sigma$; $\{n^j\}$ are the coordinates of $\mathbf{n}(P)$. Thus, we obtain a tensor field of type $(1, 1)$: at each point P there exists a linear operator Q_P smoothly dependent on the point. This field defines the stress

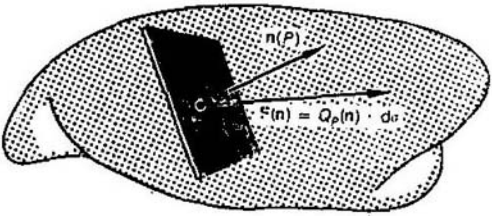

Figure 5.2

tensor. Stresses depend on the deformation of a medium, and the character of this dependence will be discussed below.

It follows from mechanical considerations (which are omitted here) that the stress tensor Q_P must be symmetric (i.e. the matrix of the operator Q_P is symmetric). The stress tensor cannot only be defined (and calculated) in a deformed elastic body. For example, this tensor does exist in an ideal liquid. It is assumed that in such a liquid there are no viscous forces, so that the force due to a stress $\mathbf{F}(\mathbf{n})$ exerted on $d\sigma$ can only be directed normal to $d\sigma$, i.e. it coincides with the force of normal pressure on $d\sigma$. Since $\mathbf{F}(\mathbf{n}) = Q_P(\mathbf{n})\, d\sigma$, the operator Q_P is diagonal, i.e.

$$Q^i_j = q(P)\, \delta^i_j; \quad F^i = Q^i_j n^j\, d\sigma = q(P)\, \delta^i_j n^j\, d\sigma = q(P)\, n^i\, d\sigma,$$
$$F^i = q(P)\, n^i(P)\, n\sigma,$$

where $q(P)$ is a scalar function called pressure at the point P. Hence, the pressure $q(P)$ does not depend on the direction $\mathbf{n}$, but depends only on the point P (the pressure may, of course, change from point to point). If in a medium $Q_P = q(P) \cdot E$, such a medium is said to obey Pascal's law. This law does not hold in every medium; for example, in a viscous liquid the stress tensor is of a more complicated form, since there exist, besides normal pressure, forces due to friction.

There is a relation between the strain and stress tensors; an analysis of this relation is one of the most important topics of elasticity

theory. In the first approximation, the stress tensor depends linearly on the strain tensor for small deformations, i.e. the following relation holds true: $Q^i_j = \alpha^{ikl}_j \eta_{kl}$, where the functions α^{ikl}_j form a tensor of rank 4 (of type (3, 1)). In $\mathbf{R}^3$ this tensor has 81 components. The linear relation between the stress and strain tensors is known as Hooke's law. Suppose the medium is homogeneous and isotropic, i.e. the form of the dependence $Q = Q(\eta)$, $Q^i_j = \alpha^{ikl}_j \eta_{kl}$ remains unchanged under orthogonal transformations in $\mathbf{R}^3$, that is, the tensor α^{ikl}_j is invariant relative to $SO(3)$. It can be demonstrated (the proof is omitted) that this leads to the following form of the dependence $Q = Q(\eta)$, $Q^i_j = Q_{ij} = \mu\eta_{ij} + \lambda \operatorname{spur} \eta \cdot \delta_{ij}$, where μ and λ are the so-called Lamé constants (they depend on the properties of the medium). The function $\Sigma \eta_{ii} = \operatorname{spur} \eta$ has a simple meaning (here we do not distinguish between the upper and lower indices, for we deal with $\mathbf{R}^3$). Consider the displacement of particles in a medium under small deformations; the functions u^i, which describe this displacement, may be looked upon as the components of the vector field defining the strain (this field is denoted by $\mathbf{v}(P)$). Then $\operatorname{spur} \eta = \sum_i \left(\frac{\partial u^i}{\partial x^i} + \frac{\partial u^i}{\partial x^i} \right) = 2 \sum_i \frac{\partial u^i}{\partial x^i} = 2 \operatorname{div} \mathbf{v}$, i.e. $\operatorname{spur} \eta$ is the divergence of the displacement field. For liquids, μ is called the viscosity coefficient.

We now turn to another class of tensor fields: the fields which can be constructed from fixed fields via algebraic operations thereby extending the given set of tensors.

5.2.2. ALGEBRAIC OPERATIONS OVER TENSORS

(1) Given two tensor fields of the same type and rank, $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ and $P^{i_1 \ldots i_p}_{j_1 \ldots j_q}$; then we can construct a new tensor field, defining it as a set of functions $C^{i_1 \ldots i_p}_{j_1 \ldots j_q} = T^{i_1 \ldots i_p}_{j_1 \ldots j_q} + P^{i_1 \ldots i_p}_{j_1 \ldots j_q}$. Using the definition of tensor, one can easily prove that $C^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ is also a tensor (verify!).

(2) If $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ is a tensor field and $f(x)$ is a smooth function on M^n, then the set $f(x) \cdot T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ is also a tensor field on M^n. The proof is straightforward.

(3) *Permutation of indices of the same type.* Given a tensor field $T_{i_1 \ldots i_q}$ (for simplicity, we consider the field of type $(0, q)$). Let us construct a new field according to the formula $P_{i_1 \ldots i_s \ldots i_\alpha \ldots i_q} = T_{i_1 \ldots i_\alpha \ldots i_s \ldots i_q}$, i.e. the operation is reduced to the permutation (renumbering) of the tensor components. Evidently, this operation is

a tensor one (verify!). This permutation interchanges two indices, i_s and i_α. Apparently, we might consider any other permutation.

Remark. Permutation of indices of different type (i.e. covariant and contravariant) is not generally a tensor operation, because while the upper indices are transformed by the matrix J, the lower indices by the matrix $(J^{-1})^T$ (see above), and these matrices are distinct.

Example. Let us consider a field of type (1, 1), i.e. the field of operators C in $T_P(M^n)$; suppose the identity $C^i_j = C^j_i$ is valid in a coordinate system (x), i.e. C is "symmetric". Here we have interchanged the upper and lower indices. In terms of the matrix C, this condition reads: $C = C^T$. Suppose the same relation $C^{i'}_{j'} = C^{j'}_{i'}$ is valid in any other coordinate system, i.e. $C' = (C')^T$. Let A be the Jacobi matrix of the transformation $(x) \to (x')$; then $C' = ACA^{-1}$, $ACA^{-1} = (ACA^{-1})^T$, $(A^TA)\,C = C\,(A^TA)$, i.e. $BC = CB$, where $B = A^TA$. Thus, our condition is equivalent to the condition that the matrices B and C are commutative, which is not always the case. If the transformation $(x) \to (x')$ had an orthogonal Jacobi matrix A, i.e. $A^TA = E$, the permutation of indices i and j would be a tensor operation.

(4) *Contraction.* Let $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ be a tensor field. Fix two indices of different type, covariant j_s and contravariant i_α, and construct the following functions: $P^{i_1\ldots i_{\alpha-1}i_{\alpha+1}\ldots i_p}_{j_1\ldots j_{s-1}j_{s+1}\ldots j_q} = \sum_i T^{i_1\ldots i_{\alpha-1},\, i_\alpha = i,\, i_{\alpha+1}\ldots i_p}_{j_1\ldots j_{s-1},\, j_s = i,\, j_{s+1}\ldots j_q}$. This is a tensor operation (verify!), it transforms a tensor of type (p, q) into a tensor of type $(p-1,\ q-1)$. If $p = q$, the initial field $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ admits complete contraction; this means that all the upper indices are contracted with all the lower indices to give a scalar function spur T which is an invariant of T, i.e. it does not change under coordinate transformations.

(5) *Tensor product.* Given two tensor fields of the general form, $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ and $P^{\alpha_1\ldots\alpha_s}_{\beta_1\ldots\beta_t}$. We can construct a new field: $C^{i_1\ldots i_p\alpha_1\ldots\alpha_s}_{j_1\ldots j_q\beta_1\ldots\beta_t} = T^{i_1\ldots i_p}_{j_1\ldots j_q}\cdot P^{\alpha_1\ldots\alpha_s}_{\beta_1\ldots\beta_t}$. Obviously, this tensor field is of type $(p+s,\ q+t)$, it is denoted by $C = T \otimes P$. Tensor product is not, in general, commutative: $T \otimes P \neq P \otimes T$.

(6) *Raising and lowering of indices.* Consider a tensor field a_{ij} of type (0, 2); let this field be non-degenerate, i.e. the matrix $A = (a_{ij})$ is non-singular at each point. Then there exists the inverse matrix A^{-1}; its coefficients are denoted by a^{ij}. Thus, $a^{ik}a_{kj} = \delta^i_j$. Given an arbitrary tensor field $T^{i_1\ldots i_p}_{j_1\ldots j_q}$, we can construct new fields

$$P^{i_2\ldots i_p}_{\alpha j_1\ldots j_q} = a_{\alpha i_1}T^{i_1 i_2\ldots i_p}_{j_1 j_2\ldots j_q} \text{ and } C^{\alpha i_1\ldots i_p}_{j_2\ldots j_q} = a^{\alpha j_1}T^{i_1\ldots i_p}_{j_1 j_2\ldots j_q}.$$

The former operation is called *index lowering* and the latter *index raising*. These are tensor operations, for they are a combination of two tensor operations: multiplication and contraction of tensors. Since the tensor field a_{ij} is non-degenerate, the operations of raising and lowering of indices are inverse. Indeed,

$$a_{j_1\alpha}C^{\alpha i_1\dots i_p}_{j_2\dots j_q} = a_{j_1\alpha}a^{\alpha s}T^{i_1\dots i_p}_{sj_2\dots j_q} = \delta^s_{j_1}T^{i_1\dots i_p}_{sj_2\dots j_q} = T^{i_1\dots i_p}_{j_1j_2\dots j_q}.$$

As a_{ij} one generally takes the metric tensor g_{ij} on a Riemannian manifold. For example, if $M^n = \mathbf{R}^n$ and $g_{ij} = \delta_{ij}$, the transformation law is the same both for upper and lower indices (in the system where $g_{ij} = \delta_{ij}$). Thus, in a Cartesian system there is no difference between upper and lower indices, and they can be raised and lowered arbitrarily; for instance, all the indices may be regarded as lower. The operations of raising and lowering of indices permit, in general, canonical identification of T_PM^n and $T^*_PM^n$. Indeed, the elements of T_PM^n are represented by vectors (tensors of type (1, 0) and the elements of $T^*_PM^n$ by covectors (tensors of type (0, 1)). Let us construct linear mappings

$$A\colon T \to T^*,\ A(\mathbf{a}) = \xi,$$

$$\text{where } \mathbf{a} \in T = T_PM^n,\ \xi \in T^* = T^*_PM^n,\ \xi_i = q_{i\alpha}a^\alpha,$$

$$B\colon T^* \to T,\ B(\eta) = \mathbf{b},\ b^i = g^{i\alpha}\eta_\alpha.$$

Then,

$$((B \circ A)\,\mathbf{a})^i = g^{i\alpha}(A\mathbf{a})_\alpha = g^{i\alpha}g_{\alpha j}a^j = \delta^i_j a^j = a^i,$$

i.e.,

$$B \circ A\colon T \to T,\ B \circ A = 1_T$$

(identity mapping). Similarly,

$$((A \circ B)\,\xi)_i = g_{i\alpha}(B\xi)^\alpha = g_{i\alpha}g^{\alpha j}\xi_j = \delta^j_i\xi_j = \xi_i,$$

i.e. $A \cdot B\colon T^* \to T^*$, $AB = 1_{T^*}$. This means that A and B are isomorphisms. Furthermore, this identification of T and T^* is invariant under coordinate transformation because only tensor operations were used. In particular, the following two diagrams are commutative:

$$\begin{tikzcd} T^* \arrow[r, "(J^{-1})^T"'] & T^* \\ T \arrow[u, "A"] \arrow[r, "J"'] & T \arrow[u, "A"'] \end{tikzcd} \qquad \begin{tikzcd} T^* \arrow[r, "(J^{-1})^T"'] \arrow[d, "B"'] & T^* \arrow[d, "B"] \\ T \arrow[r, "J"'] & T \end{tikzcd}$$

(7) *Symmetrization*. Let us first consider a pair of indices of the same type and, using the tensor field $T_{\dots i \dots j \dots}$, construct

a new field $P_{\ldots i \ldots j \ldots} = T_{\ldots j \ldots i \ldots}$ (by interchanging these indices). We define general symmetrization as follows: $P_{j_1 \ldots j_q} = T_{(j_1 \ldots j_q)} = \frac{1}{q!} \sum_{(\sigma)} T_{\sigma(j_1 \ldots j_q)}$, where summation extends over all permutations of the indices $i_1, \ldots, i_q$. Symmetrization is a tensor operation (verify!).

Definition. A field $T^{i_1 \ldots i_p}_{j_1 \ldots j_p}$ is called *symmetric* if it does not change under the permutation of any two indices of the same type.

Symmetrization of a symmetric field results in the same field. Example: the metric tensor g_{ij} is symmetric.

(8) *Alternation*. Let us consider a pair of neighbouring indices of the same type and, using the tensor field $T_{\ldots ij}$, construct a new field $P_{\ldots ij \ldots} = \frac{1}{2}(-T_{\ldots ji \ldots} + T_{\ldots ij \ldots})$. Alternation is defined as follows:

$$P_{j_1 \ldots j_q} = T_{[j_1 \ldots j_q]} = \frac{1}{q!} \sum_{(\sigma)} (-1)^{\varphi(\sigma)} T_{\sigma(j_1 \ldots j_q)}.$$

Here $\varphi(\sigma)$ is the parity of the permutation σ (the notation $(-1)^\sigma$ is sometimes used). Alternation is a tensor operation (verify!).

Definition. *A tensor field* $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ is called *skew-symmetric* if its components change the sign under the transposition of any two neighbouring indices of the same type.

Lemma 1. *Alternation does not change a skew-symmetric tensor* $T_{i_1 \ldots i_k}$ *(for simplicity, we consider tensors with covariant indices). Alternation makes a symmetric tensor vanish.*

The *proof* directly follows from the definition of a symmetric and a skew-symmetric tensor.

Here we end the discussion of the main algebraic operations over tensors and turn to the concept of a symmetric and a skew-symmetric operator. The example presented above shows that the attempt at defining, say, a symmetric operator $C = (c^i_j)$ by the formula $C = C^T$ (which must hold in all coordinate systems) fails, for this definition is not a tensor one, and the relation $C = C^T$ becomes invalid under coordinate transformations. The concept of symmetry and skew symmetry of an operator is valid if only the Riemannian metric exists on the manifold (or, in a more general form, if there exists a nondegenate tensor field of type (0, 2) or (2, 0)). Let us consider M^n with the metric g_{ij} which generates in each $T_P M^n$ the scalar product

$$\langle \mathbf{a}, \mathbf{b} \rangle_P = g_{ij}(P) a^i b^j, \quad \mathbf{a}, \mathbf{b} \in T_P M^n.$$

Definition. An operator T (at one point P) or an operator field of type (1, 1) is called *symmetric* (or *skew-symmetric*) if the identity $\langle T\mathbf{a}, \mathbf{b}\rangle \equiv \langle \mathbf{a}, T\mathbf{b}\rangle$ holds for any $\mathbf{a}, \mathbf{b} \in T_PM^n$ at this point (or at all points for the operator field); in the skew-symmetric case the corresponding identity is $\langle T\mathbf{a}, \mathbf{b}\rangle \equiv -\langle \mathbf{a}, T\mathbf{b}\rangle$, $\mathbf{a}, \mathbf{b} \in T_P M^n$.

Explanation (say, for a symmetric operator): $\langle \mathbf{a}, T\mathbf{b}\rangle = \langle T\mathbf{a}, \mathbf{b}\rangle$, $g_{ij}a^i (T\mathbf{b})^j = g_{ij}T^j_k b^k = (g_{ij}T^j_k)\, a^i b^k = T_{ik}a^i b^k$, where $T_{ik} = g_{ij}T^j_k$. Furthermore, $\langle T\mathbf{a}, \mathbf{b}\rangle = g_{ij}\,(T\mathbf{a})^i b^j = g_{ij}T^i_k a^k b^j = g_{jk}T^j_i a^i b^k = T_{ki}a^i b^k$, where $T_{ki} = g_{ik}T^j_i$. Since $T_{ik}a^i b^k \equiv T_{ki}a^i b^k$, we have $T_{ik} = T_{ki}$. Thus, after index lowering in a tensor T of type (1, 1) the symmetry condition (imposed on a tensor of type (0, 2)) takes the usual form, i.e. T_{ik} does not change under index transposition. That T_{ik} is skew-symmetric after index lowering (in the skew-symmetric case) can be demonstrated in a similar way.

5.2.3. SKEW-SYMMETRIC TENSORS

Here we shall consider covariant skew-symmetric tensors $T_{i_1\ldots i_k}$ (or tensor fields).

Lemma 2. *A skew-symmetric tensor $T_{i_1\ldots i_n}$ of maximal rank n on M^n is completely defined by only one component $T_{12\ldots n}$ (the so-called principal component); the other components differ from the principal component by the factor $(-1)^\sigma$, i.e.*

$$T_{i_1\ldots i_n} = (-1)^\sigma T_{12\ldots n}\ (i_1, \ldots, i_n) = \sigma\,(1, 2, \ldots, n).$$

Proof. Let us fix a coordinate system $x^1, \ldots, x^n$, then in this system $T_{i_1\ldots i_n} = (-1)^\sigma T_{12\ldots n}$ (see the definition of a skew-symmetric tensor). After the transformation $(x) \to (x')$ we have

$$T_{1'\ldots n'} = \frac{\partial x^{i_1}}{\partial x^{1'}} \cdots \frac{\partial x^{i_n}}{\partial x^{n'}}\, T_{i_1\ldots i_n}$$

$$= \Big(\sum_\sigma (-1)^\sigma \frac{\partial x^{i_1}}{\partial x^{1'}} \cdots \frac{\partial x^{i_n}}{\partial x^{n'}}\Big)\, T_{12\ldots n} = (\det J)\cdot T_{12\ldots n}.$$

Thus, to describe a skew-symmetric tensor of maximal rank it is sufficient to know its principal component (in one coordinate system because the principal components in other coordinate systems are obtained via multiplication by the transformation Jacobian). The lemma is proved.

Skew-symmetric tensors admit an important operation called *exterior multiplication*. Let $T_{i_1\ldots i_k}$ and $P_{j_1\ldots j_q}$ be two skew-symmetric

tensors. Define a new skew-symmetric tensor (denoted as $R = T \wedge P$) by the relation

$$R_{i_1\ldots i_k j_1\ldots j_h} = T_{[i_1\ldots i_k} P_{j_1\ldots j_q]} = \frac{1}{k!q!} \sum_{\sigma} \left(\sum_{\sigma} (-1)^{\sigma} T_{\sigma(i_1\ldots i_k} P_{j_1\ldots j_q)} \right).$$

This multiplication is a bilinear operation; the ranks of the tensors are added. Let us interpret the operation in terms of exterior differential forms. First consider a skew-symmetric tensor $T_{i_1\ldots i_k}$ given at one point on M^n. Then $T_{i_1\ldots i_k}$ defines a skew-symmetric multilinear mapping T: $(T_* \times \ldots \times T_*)(k \text{ times}) \to \mathbf{R}$, where $T_* = T_P M^n$. Fix a coordinate system $x^1, \ldots, x^n$ and consider these coordinates as smooth functions in a neighbourhood of a point P. Then the differentials $dx^1, \ldots, dx^n$ of these functions are valid. The differential df of any smooth function $f(x)$ is an element of the space adjoint to T_*, i.e. a linear functional on T_*. Indeed, if $\mathbf{a} \in T_*$, then

$$\frac{df}{d\mathbf{a}} = \frac{df(\gamma(t))}{dt}\Big|_{t=0},$$

where $\gamma(0) = P$, $\dot{\gamma}(0) = \mathbf{a}$, i.e.

$$\frac{df}{d\mathbf{a}} = \frac{\partial f}{\partial x^k}\frac{dx^k}{dt}\Big|_{t=0} = \frac{\partial f}{\partial x^k} a^k. \text{ Thus, } \frac{df}{dt} = \frac{\partial f}{\partial x^k} a^k,$$

i.e. if $f = x^i$, then $\frac{dx^i}{dt} = \frac{\partial x^i}{\partial x^k} a^k = \delta^i_k a^k = a^i$; $dx^k = a^k\,dt$.

Hence, dx^k may be looked upon as functionals on T_*, i.e. elements of T_*. Assuming dt small, we may consider it as a proportionality coefficient, which is omitted below. If $\mathbf{e}_1, \ldots, \mathbf{e}_n \in T_*$ is a basis in T_*, then for basis in T^* we take $dx^1, \ldots, dx^n$ assuming that $dx^k(\mathbf{e}_\alpha) = \delta^k_\alpha$, i.e. $dx^k(\mathbf{a}) = a^k$. Let $\Lambda(dx^1, \ldots, dx^n)$ be an exterior algebra with the generators $dx^1, \ldots, dx^n$ satisfying the relations $dx^i \wedge dx^j + dx^j \wedge dx^i = 0$, $i \neq j$. The operation Λ is bilinear, and the algebra Λ is generated (in the additive sense) by the monomials $dx^1, \ldots, dx^n$

$$dx^i \wedge dx^j, \quad i < j; \ldots; dx^{i_1} \wedge \ldots \wedge dx^{j_k},$$
$$i_1 < \ldots < i_k,\ dx^1 \wedge \ldots \wedge dx^n.$$

As is known from algebra, these monomials do not obey linear relations with constant coefficients; the multiplicative generators of the algebra $\Lambda(dx^1, \ldots, dx^k)$ are represented by $dx^1, \ldots, dx^k$, i.e. by the basis in T^*. The algebra Λ can be decomposed into the direct sum of linear subspaces Λ^k: $\Lambda = \bigoplus_{k=0}^{n} \Lambda^k$, where Λ^k, $1 \leqslant k \leqslant n$, are generated by the monomials $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, $i_1 < \ldots < i_k$;

and $\Lambda^0 \cong \mathbf{R}$ is generated by unity element 1. The dimension of the algebra Λ is equal to 2^n.

Let us consider a new algebra $\Lambda(M^n)$ whose elements are linear combinations $\omega^{(k)} = T_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ and all possible linear combinations $\sum_{k=0}^{n} \omega^{(k)}$, where $T_{i_1 \ldots i_k}(x)$ is a skew-symmetric tensor field of rank k, and the indices $i_1 \ldots i_k$ are arranged in increasing order. Multiplication in $\Lambda(M^n)$ is defined below.

We have defined the element $\omega^{(k)}$ in a given coordinate system $x^1, \ldots, x^n$, $0 \leqslant k \leqslant n$. What does it happen to this object under the transformation $(x) \to (x')$?

Lemma 3. *The elements* $\omega^{(k)} = T_{i_1 \ldots i_k}(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ *are defined invariantly, i.e. under the transformation* $(x) \to (x')$ *we have*

$$T_{i'_1 \ldots i'_k} dx^{i'_1} \wedge \ldots \wedge dx^{i'_k} \equiv T_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}.$$

Proof. It is sufficient to make the substitution $(x) \to (x')$ and use the transformation laws for $T_{i_1 \ldots i_k}(x)$ and dx^i.

Definition. The elements $\omega^{(k)}$ of the algebra $\Lambda(M^n)$ are called *exterior differential forms*.

The algebra $\Lambda(M^n)$ is infinite-dimensional. Each exterior form uniquely defines the collection of components $T_{i_1 \ldots i_k}$ of a skew-symmetric tensor; the converse is also true: any skew-symmetric tensor (a tensor field, to be more precise) uniquely defines an exterior form. The apparatus of exterior forms is one of the apparatuses to describe skew-symmetric fields.

Remark. While defining an exterior form, we might consider linear combinations $A_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, where $A_{i_1 \ldots i_k}$ is an arbitrary tensor field (not necessarily skew-symmetric) and summation extends over all (not ordered) collections $i_1, \ldots, i_k$. Since the generators $dx^1, \ldots, dx^n$ are skew-symmetric with respect to exterior multiplication, we have

$$A_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k} = A_{[i_1 \ldots i_k]} dx^{i_1} \wedge \ldots \wedge dx^{i_k},\ i_1 < \ldots < i_k,$$

i.e. any combination $A_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ generates an exterior differential form.

Let us define multiplication in $\Lambda(M^n)$. If $\omega^{(k)} = T_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ $(i_1 < \ldots < i_k)$ and $\omega^{(s)} = P_{j_1 \ldots j_s} dx^{j_1} \wedge \ldots \wedge dx^{j_s}$ $(j_1 < \ldots < j_s)$ are exterior forms, their product $\omega^{(k+s)} = \omega^{(k)} \wedge \omega^{(s)}$ is

defined as the form

$$\begin{aligned}\omega^{(k+s)} &= (T_{i_1\dots i_k}\,dx^{i_1}\wedge\dots\wedge dx^{i_k})\wedge(P_{j_1\dots j_s}\,dx^{j_1}\wedge\dots\wedge dx^{j_s})\\ &= T_{i_1\dots i_k}P_{j_1\dots j_s}\,dx^{i_1}\wedge\dots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\dots\wedge dx^{j_s}\\ &\qquad\qquad (i_1<\dots<i_k)\quad (j_1<\dots<j_s)\\ &= T_{[i_1\dots i_k}P_{j_1\dots j_s]}\,dx^{i_1}\wedge\dots\wedge dx^{i_{k+s}}.\\ &\qquad\qquad (i_1<\dots<i_{k+s}).\end{aligned}$$

Apparently, multiplication of forms coincides with exterior multiplication of the corresponding tensor fields. Thus, $\Lambda(M^n)$ is provided with the structure of an algebra with unity element, and multiplication is associative (because multiplication in the algebra $\Lambda(dx^1, \dots, dx^n)$ is associative), but not commutative. The algebra $\Lambda(M^n)$ is decomposed into the direct sum of linear subspaces $\Lambda^{(k)}(M^n) = \{T_{i_1\dots i_k}\,dx^{i_1}\wedge\dots\wedge dx^{j_k}\}$. The number k is called the degree of the form $\omega^{(k)}$. The elements of $\omega^{(k)}$ are called homogeneous elements of the algebra $\Lambda(M^n)$. It is convenient to interpret differential forms $\omega^{(k)}$ as multilinear skew-symmetric mappings. If $\mathbf{a}_1, \dots, \mathbf{a}_k \in T_*M^n$, then, assuming $dx^i(\mathbf{a}) = a^i$ (the ith coordinate), i.e. omitting the factor dt, we obtain

$$\begin{aligned}&\omega^{(k)}(\mathbf{a}_1, \dots, \mathbf{a}_k)\\ &\qquad = (T_{i_1\dots i_k}(x)\,dx^{i_1}\wedge\dots\wedge dx^{i_k})(\mathbf{a}_1, \dots, \mathbf{a}_k)\\ &\qquad = T_{i_1\dots i_k}(x)\,dx^{[i_1}(\mathbf{a}_1)\cdot\,\dots\,\cdot dx^{i_k]}(\mathbf{a}_k)\\ &\qquad = T_{i_1\dots i_k}(x)\,a_1^{[i_1}\cdot\,\dots\,\cdot a_k^{i_k]}\\ &\qquad = T_{[i_1\dots i_k]}a_1^{i_1}\dots a_k^{i_k} = T_{i_1\dots i_k}a_1^{i_1}\dots a_k^{i_k}.\end{aligned}$$

Thus

$$\omega^{(k)}: \underbrace{T_x\times\dots\times T_x}_{k} \to R,$$

$$\omega^{(k)}(\mathbf{a}_{\sigma(1,}\ a_2\dots a_{k)}) = (-1)^{\sigma}\,\omega^{(k)}(\mathbf{a}_1, \dots, \mathbf{a}_k).$$

Let $x^1, \dots, x^n$ be Cartesian coordinates in $\mathbf{R}^n$. Let also $\omega^{(n)} = dx^1\wedge\dots\wedge dx^n$ be an exterior form of maximal rank and let $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbf{R}^n$ be an arbitrary collection of vector arguments in $\mathbf{R}^n$. We write vol $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$ for the n-dimensional volume of the parallelepiped $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$ spanned by $\mathbf{a}_1, \dots, \mathbf{a}_n$ (see Fig. 5.3).

Lemma 4. *Let* $A = \begin{pmatrix} a_1^1 & \dots & a_1^n \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ a_n^1 & \dots & a_n^n \end{pmatrix}$ *be a matrix constructed of the coordinates of* $\mathbf{a}_1, \ldots, \mathbf{a}_n$. *Then* $\det A = \operatorname{vol} \Pi (\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Proof. For $n = 1$ the statement is evident. Suppose the formula is proved for $k \leqslant n - 1$. Since $\operatorname{vol} \Pi (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ remains unchanged under rotations, the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$ can be arranged so as to

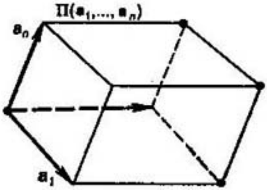

Figure 5.3

lie in the plane spanned by $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$, where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis in $\mathbf{R}^n$. Then the matrix A takes the form

$$A' = \left|\begin{array}{c|c} B' & * \\ \hline 0 \ \ldots \ldots \ 0 & a_n'^n \end{array}\right|.$$

Hence,

$$\begin{aligned} \det A = \det A' &= (\det B') \cdot (a')_n^n \\ &= (\operatorname{vol} \Pi (\mathbf{a}_1, \ldots, \mathbf{a}_{n-1})) \cdot (a')_n^n = \operatorname{vol} \Pi (\mathbf{a}_1, \ldots, a_n) \end{aligned}$$

(see Fig. 5.4). Here $(a')_n^n$ is the projection of $\mathbf{a}_n$ onto $\mathbf{e}_n$. The lemma is proved.

Lemma 5. *Let* $\mathbf{R}^n$ *be referred to Cartesian coordinates* $x^1, \ldots, x^n$ *and let* $\omega^{(n)} = dx^1 \wedge \ldots \wedge dx^n$ *be a form of maximal rank. Then* $\omega^{(n)} (\mathbf{a}_1, \ldots, \mathbf{a}_n) = \operatorname{vol} \Pi (\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Proof. Since $dx^i (\mathbf{a}) = a^i$, we have

$$\begin{aligned} \omega^{(n)} (\mathbf{a}_1, \ldots, \mathbf{a}_n) &= (dx^1 \wedge \ldots \wedge dx^n) (\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ &= dx^{[1} (\mathbf{a}_1) \cdot \ldots \cdot dx^{n]} (\mathbf{a}_n) \\ &= a_1^{[i_1} \cdot \ldots \cdot a_n^{i_n]} = \det A, \end{aligned}$$

where $A = \begin{pmatrix} a_1^1 & \dots & a_1^n \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ a_n^1 & \dots & a_n^n \end{pmatrix}$, i.e. $\det A = \operatorname{vol} \Pi (\mathbf{a}_1, \ldots, \mathbf{a}_n)$. The lemma is proved.

Thus, the value of the standard form $dx^1 \wedge \ldots \wedge dx^n$ on any collection $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is equal to the Euclidean volume of the parallelepiped spanned by $\mathbf{a}_1, \ldots, \mathbf{a}_n$. In particular, the volume of a parallelepiped (in such an interpretation) is not a scalar; for instance, it changes the sign under an odd permutation of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ (it is an "oriented volume"). Thus, we have represented $\omega = dx^1 \wedge \ldots \wedge dx^n$ as an n-linear functional of $\mathbf{a}_1, \ldots, \mathbf{a}_n$. This interpretation of volume as the value of an exterior form (of the arguments $\mathbf{a}_1, \ldots, \mathbf{a}_n$)

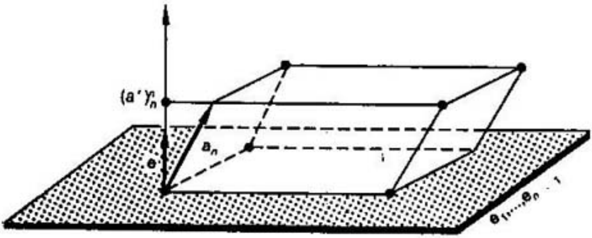

Figure 5.4

is rather fruitful; it leads to many interesting geometric facts. For example, such an interpretation implies the invariance of a multiple integral under a change of variables.

A similar interpretation is possible with forms of a lower degree. Let, for example, a form $\omega^{(k)} = dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ be given in $\mathbf{R}^n$.

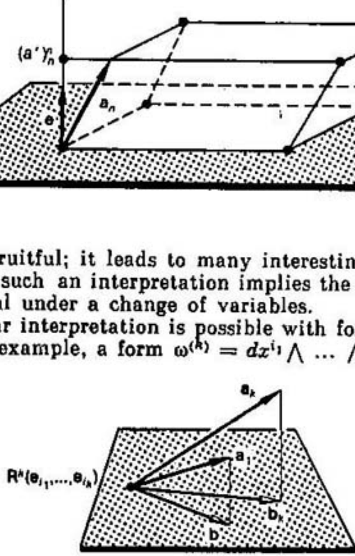

Figure 5.5

It is required to find the value of $(dx^{i_1} \wedge \ldots \wedge dx^{i_k})(\mathbf{a}_1, \ldots, \mathbf{a}_k)$, where $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is an arbitrary set of vectors in $\mathbf{R}^n$. Exercise: prove that $\omega^{(k)}(\mathbf{a}_1, \ldots, \mathbf{a}_k) = \operatorname{vol} \Pi(\mathbf{b}_1, \ldots, \mathbf{b}_k)$, where $\operatorname{vol} \Pi(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ is the volume of the k-dimensional parallelepiped located in the coordinate plane $\mathbf{R}^k(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k})$ and spanned by the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbf{R}^k$ which are orthogonal projections of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ on the plane $\mathbf{R}^k(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k})$ (see Fig. 5.5.).

If the form $\omega^{(k)}$ is the combination $\omega_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, its value on $\mathbf{a}_1, \ldots, \mathbf{a}_k$ will be the combination of the volumes $\operatorname{vol} \Pi (\mathbf{b}_1, \ldots, \mathbf{b}_k)$ with weights $\omega_{i_1 \ldots i_k}$.

There is a relation between exterior form of maximal degree and the volume of a bounded domain on a Riemannian manifold. The reader is supposed to be familiar with the definition of a multidimensional (multiple) Riemann integral. Let M^n be a smooth Riemannian manifold and let D be an open domain in M^n for which there exists a diffeomorphism onto an open bounded domain U in $\mathbf{R}^n$ referred to Cartesian coordinates $x^1, \ldots, x^n$. Of course, this diffeomorphism exists not for every domain D on M^n, but for simplicity we shall confine ourselves to "sufficiently small" domains. Let $g_{ij}(x)$ be a metric and let $g(x)$ be the determinant of the matrix $(g_{ij}(x))$.

Definition. The *volume of a domain* $D \subset M^n$ is the number

$$V(D) = \operatorname{vol} D = \int \underset{U(x)}{\cdots} \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n,$$

where $x^1, \ldots, x^n$ are Cartesian coordinates in the domain $U(x) \subset \mathbf{R}^n$ (see Fig. 5.6).

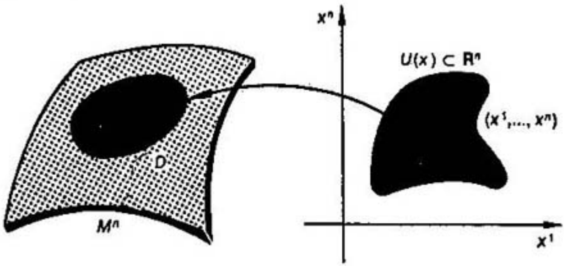

Figure 5.6

This definition needs substantiation, i.e. it is required to demonstrate that in simple cases this formula gives the same values of the volume as those obtained from other considerations.

(1) Let $M^n = \mathbf{R}^n$ and let $U = D \subset \mathbf{R}^n$ be a bounded domain in $\mathbf{R}^n$, $g_{ij} = \delta_{ij}$, i.e. $g(x) \equiv 1$, so that

$$\operatorname{vol} D = \int \underset{U(x)}{\cdots} \int dx^1 \ldots dx^n,$$

which coincides with the customary "Euclidean" definition of the volume of a bounded domain in $\mathbf{R}^n$. It is assumed here that the symbol $dx^1, \ldots, dx^n$ implies, in addition to its formal meaning (see

the definition of the Riemann integral), also the following: if $dx^1, \ldots, dx^n$ are considered as infinitesimal quantities, $d\sigma = dx^1 \ldots dx^n$ is the infinitesimal volume of an infinitesimal parallelepiped with edges, $dx^1 \ldots, dx^n$ measured in Cartesian coordinates $x^1, \ldots, x^n$. Thus, the volume of a bounded domain $U(x) \subset \mathbf{R}^n$ can be represented as the sum of infinitely many volumes of infinitesimal parallelepipeds calculated by the ordinary formula (the volume of a rectangular parallelepiped is equal to the product of all its edges emerging from one vertex).

(2) Let M^n be a smooth submanifold in $\mathbf{R}^N$ and let $g_{ij}(x)$ be the induced Riemannian metric on $M^n \subset \mathbf{R}^N$ (the ambient metric is Euclidean). Also, let $x^1, \ldots, x^n$ be curvilinear coordinates in the neighbourhood of a point P on M^n and let a domain D lie in this neighbourhood. Then we may assume that the volume of the domain D, vol D, is the sum of infinitely many volumes of infinitesimal parallelepipeds (which are no longer rectangular) $\{\Pi_k\}$ obtained when D is subdivided by the coordinate planes $x^i = \text{const}$ $(1 \leqslant i \leqslant n)$ (with "infinitesimal spacing") (see Fig. 5.7). Let us consider an infinitesimal "curvilinear" parallelepiped Π_k.

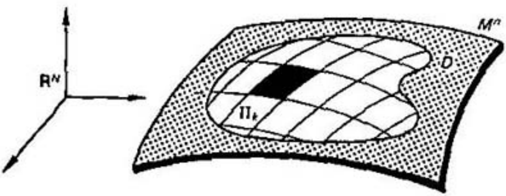

Figure 5.7

Since M^n is a smooth submanifold in $\mathbf{R}^N$, we may assume that Π_k is well approximated by the "linear" parallelepiped $\widetilde{\Pi}_k$ contained in $T_P M^n$, where P is a vertex of Π_k

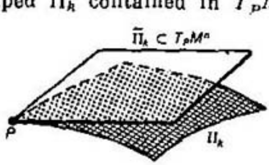

Figure 5.8

(see Fig. 5.8). The sides of Π_k are directed along the coordinate lines $\{x^i\}$ through P. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be the velocity vectors of these lines

(i.e. $\mathbf{a}_i$ is the velocity vector of the line along which only the coordinate x^i changes, while all the others are fixed). Since g_{ij} is induced by the Euclidean metric, vol Π_k = vol $\widetilde{\Pi}_k$ coincides with the Euclidean (n-dimensional) volume of $\widetilde{\Pi}_k$ imbedded in $\mathbf{R}^n$. Thus, vol $D \cong \sum_{(i)}$ vol $\widetilde{\Pi}_k$. Let at P there be given, in addition to curvilinear coordinates $x^1, \ldots, x^n$, orthogonal coordinates $y^1, \ldots, y^n$; this means that we define Cartesian coordinates in T_PM^n. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be unit velocity vectors of these new coordinate lines at the point P. Then in the coordinates $y^1, \ldots, y^n$ we have $g_{ij}(y) = \delta_{ij}$ at P (generally, only at one point). Let $(x) \to (y)$ be a coordinate transformation and J the Jacobi matrix at P. The matrix J,

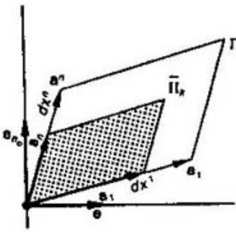

Figure 5.9

operating in T_PM^n, sends $\mathbf{e}_1, \ldots, \mathbf{e}_n$ into $\mathbf{a}_1, \ldots, \mathbf{a}_n$. We now prove that vol $\widetilde{\Pi}_k = \sqrt{g(x)}\, d\sigma^n$, where $d\sigma^n = dx^1, \ldots, dx^n$ is the Euclidean volume of a rectangular parallelepiped with sides $dx^1, \ldots, dx^n$. Consider a parallelepiped Π_k spanned by $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Then vol $\Pi'_k(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \sqrt{g(x)}$, where vol denotes the Euclidean volume. Indeed, since in the coordinates $y^1, \ldots, y^n$ the tensor g_{ij} is of the form $g_{ij}(y) = \delta_{ij}$, we have $(g_{ij}(x)) = J(g_{ij}(y))J^T = JEJ^T = JJ^T$, i.e. $g(x) = \det(g_{ij}(x)) = \det(JJ^T) = (\det J)^2$, and hence $\det J = \sqrt{g(x)}$. Since J transforms the orthogonal coordinate frame $\mathbf{e}_1, \ldots, \mathbf{e}_n$ into $\mathbf{a}_1, \ldots, \mathbf{a}_n$, we have, according to Lemma 4, $\sqrt{g(x)} = \det J =$ vol $\Pi'_k(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. The statement is proved.

Since $\widetilde{\Pi}_i$ is spanned by $(dx^1)\,\mathbf{a}_1, \ldots, (dx^n)\,\mathbf{a}_n$ (see Fig. 5.9),

$$\text{vol}\, \widetilde{\Pi}_k(\mathbf{a}_1 \cdot dx^1, \ldots, \mathbf{a}_n \cdot dx^n)$$
$$= (\text{vol}\, \Pi'_k(\mathbf{a}_1, \ldots, \mathbf{a}_n)) \cdot dx^1 \cdot \ldots \cdot dx^n = \sqrt{g(x)}\, dx^1 \cdot \ldots\ dx^n.$$

Because the volume of the domain D consists of the volumes $\tilde{\Pi}_k$, we obtain $\mathrm{vol}\, D = \int \underset{U(x)}{\cdots} \int \sqrt{g(x)}\, dx^1 \ldots dx^n$, which is what was required.

(3) Let $M^2 \subset \mathbf{R}^3$ be a smooth submanifold given by the radius vector $\mathbf{r} = \mathbf{r}(u, v)$, then by the definition of the area of a domain D on M^2 we have

$$\mathrm{vol}\, D = \iint_{U(u,\, v)} \sqrt{g(u, v)}\, du\, dv, \quad g = \det \begin{pmatrix} E & F \\ F & G \end{pmatrix} = EG - F^2,$$

i.e.

$$\mathrm{vol}\, D = \int \int \sqrt{FG - F^2}\, du\, dv = \iint_{U(u,\, v)} |[\mathbf{r}_u, \mathbf{r}_v]|\, du\, dv,$$

where $[\ ,\ ]$ stands for the vector product and $|[\ ,\ ]|$ for its magnitude. Thus, the formula for the area of a domain coincides with one of the formulas of classical analysis.

(4) Let a submanifold $M^{n-1} \subset \mathbf{R}^n$ be given as the graph of a smooth function $x^n = f(x^1, \ldots, x^{n-1})$. Find the volume of a bounded

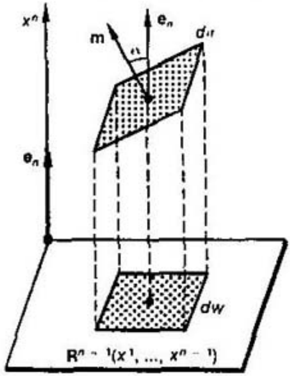

Figure 5.10

domain $D \subset M^{n-1}$. We may assume that $\mathrm{vol}\, D = \Sigma\, (d\sigma)$, where $d\sigma$ is the infinitesimal volume of an infinitesimal parallelepiped. Projecting this parallelepiped onto the hyperplane $\mathbf{R}^{n-1}(x^1, \ldots, \ldots, x^{n-1})$, we obtain the infinitesimal volume $d\omega = dx^1 \ldots dx^n$ related to $d\sigma$ by $d\sigma = \dfrac{d\omega}{\cos \alpha} = \dfrac{1}{\cos \alpha} \cdot dx^1 \ldots dx^{n-1}$ (see Fig. 5.10).

Here α is the angle between the normal $\mathbf{m}$ to M^{n-1} and the vector $\mathbf{e}_n$. Let us find $\cos\alpha$. Evidently, $\cos\alpha = \frac{\langle \mathbf{e}_n \cdot \mathbf{m}\rangle}{|\mathbf{e}_n|\cdot|\mathbf{m}|} = \langle \mathbf{e}_n, \mathbf{m}\rangle$, where

$$\mathbf{e}_n = (0, \ldots, 0, 1), \quad \mathbf{m} = \frac{\operatorname{grad} F}{|\operatorname{grad} F|},$$

and

$$F(x^1, \ldots, x^n) = x^n - f(x^1, \ldots, x^{n-1}),$$
$$\operatorname{grad} F(-f_{x^1}, \ldots, -f_{x^{n-1}}, 1),$$
$$|\operatorname{grad} F| = \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}, \quad \langle \mathbf{e}_n, \mathbf{m}\rangle = \frac{1}{\sqrt{1+\sum_{i=1}^{n-1} (f_{x^i})^2}},$$
$$d\sigma = \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}\, dx^1 \cdot \ldots \cdot dx^{n-1}.$$

Thus

$$\operatorname{vol} D = \int_{U(x^1, \ldots, x^{n-1})} \!\!\cdots\! \int \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}\, dx^1 \ldots dx^{n-1}.$$

Does this formula coincide with the general formula suggested above for calculating the volume of a domain? Let us find the induced metric on M^{n-1} in the explicit form. We have

$$ds^2 = (1 + (f_{x^i})^2 (dx^i)^2 + 2f_{x^i} f_{x^j}\, dx^i\, dx^j,$$
$$(g_{ij}) = \begin{pmatrix} 1+(f_{x^1})^2 & & f_{x^i} f_{x^j} \\ & \ddots & \\ f_{x^i} f_{x^j} & & 1+(f_{x^n})^2 \end{pmatrix},$$

$g(x) = 1 + \sum_{i=1}^{n-1} f_{x^i}^2$ (verify!). For example, if $n = 3$, then $g(x) = 1 + f_x^2 + f_y^2$. *Problem*: calculate the area of a circle of radius r on a Lobachevskian plane and on a two-dimensional sphere.

Thus, we have given several evidences for the general formula $\operatorname{vol} D = \int \underset{U(x)}{\cdots} \int \sqrt{g(x)}\, dx^1 \ldots dx^n$ which permits the calculation of the volume of a bounded domain D on a Riemannian manifold; we have assumed, however, that D is contained in a single chart.

While deriving the general formula, we have proved that the number $\operatorname{vol} D$ can be represented as

$$\operatorname{vol} D = \int \underset{U(x)}{\cdots} \int \sqrt{g(x)} \cdot (dx^1 \wedge \ldots \wedge dx^n)(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n),$$

i.e.

$$(dx^1 \wedge \ldots \wedge dx^n)(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n)$$
$$= dx^{[1}(\mathbf{a}_1 dt^1) \ldots dx^{n]}(\mathbf{a}_n dt^n) = dt^{[i_1} \ldots dt^{i_n]} = dt^1 \cdot \ldots \cdot dt^n,$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are tangent velocity vectors to the coordinate lines $\{x^i\}$ at the point P and $dt^1, \ldots, dt^n$ are infinitesimal displacements from P along these lines; these displacements form the infinitesimal parallelepiped $\widetilde{\Pi}_k(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n)$. The formula for the volume can, therefore, be written as

$$\operatorname{vol} D = \int \underset{U(x(t))}{\cdots} \int \sqrt{g(x(t))}\,(dx^1 \wedge \ldots \wedge dx^n)(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n),$$

or in a compact form

$$\int \underset{U(x)}{\cdots} \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n.$$

Apparently, the integral $\int \underset{U(x)}{\cdots} \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n$ does not depend on the choice of the local coordinate system. Indeed for the transformation $(x) \to (y)$ we have

$$\int \underset{U(y)}{\cdots} \int \sqrt{g(y)}\, dy^1 \wedge \ldots \wedge dy^n$$
$$= \int \underset{U(y)}{\cdots} \int (\det J_y) \cdot dy^1 \wedge \ldots \wedge dy^n$$
$$= \int \underset{U(y(x))}{\cdots} \int (\det J_x) \cdot \det\left(\frac{\partial x}{\partial y}\right) \cdot dy^1(x) \wedge \ldots \wedge dy^n(x)$$
$$= \int \underset{U(y(x))}{\cdots} \int (\det J_x) \det\left(\frac{\partial x}{\partial y}\right) \cdot \det\left(\frac{\partial y}{\partial x}\right) \cdot dx^1 \wedge \ldots \wedge dx^n$$
$$= \int \underset{U(y(x))}{\cdots} \int \det J_x\, dx^1 \wedge \ldots \wedge dx^n$$
$$= \int \underset{U(y(x))}{\cdots} \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n.$$

It is sometimes convenient to consider the integral

$$\int \underset{D}{\cdots} \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n$$

as an "exterior differential form" such that its application to the family of vectors $\mathbf{a}_1(x)\,dt^1, \ldots, \mathbf{a}_n(x)\,dt^n$ gives a number, the volume of domain D. Let us consider another structure related to skew-symmetric tensors of rank 2.

Definition. A *skew-symmetric scalar product* is said to be defined on M^{2n} if there is given on M^{2n} an exterior 2-form $\omega^{(2)} = \omega_{ij}(x)\,dx^i \wedge dx^j$ whose matrix $(\omega_{ij}(x))$ is non-singular, i.e. $\det \omega_{ij} \neq 0$. (Such 2-forms are called non-singular.)

The non-singularity of ω_{ij} implies the existence of the inverse field, ω^{ij}, i.e. $\omega_{ij}\omega^{jk} = \delta^k_i$. The form of maximal rank, Ω, can be associated with the form $\omega^{(2)}$: namely, $\Omega = \Omega^{(2n)} = \omega \times \ldots \times \omega$ (n times).

Theorem 1. *The skew-symmetric product* $\omega^{(2)}: \omega^{(2)}(\mathbf{a}, \mathbf{b}) = \omega_{ij}a^ib^j$ *is non-singular at a point* $P \in M^{2n}$ *if and only if the form* $\Omega^{(2n)}$ *does not vanish at this point.*

Proof. We proceed by analogy with the proof of the formula $d\sigma = \sqrt{g}\,dx^1 \wedge \ldots \wedge dx^n$ for the Riemannian metric. The proof of Theorem 1 is thus reduced to verifying a similar formula for the skew-symmetric scalar product. Since Ω is the form of maximal rank, in the coordinates $x^1, \ldots, x^n$ it becomes $\Omega = f(x)\,dx^1 \wedge \ldots \wedge dx^n$, where $f(x)$ is a smooth function on M^{2n}. That Ω does not vanish at the point P means that $f(x) \neq 0$. If we demonstrate that $(f(x) = \sqrt{\det \omega_{ij}}$, the theorem will be proved. Let us consider the matrix (ω_{ij}) at P. In the neighbourhood of P we can choose new corodinates $y^1, \ldots, y^n$ in which $(\omega_{ij}(y))$ (at the point P) takes the form

$$\left(\begin{array}{cc|ccccc}
0 & 1 & & & 0 & & \\
-1 & 0 & & & & & \\
\hline
& & \ddots & & & & \\
& & & 0 & 1 & & \\
& 0 & & -1 & 0 & & \\
& & & & & 0 & \\
& & & & & & \ddots \\
& & & & & & & 0
\end{array}\right).$$

We recall that any skew-symmetric matrix can be reduced to this form (irrespective of whether it is singular or non-singular). If $(\omega_{ij}(y))$ is non-singular, it does not have zero eigenvalue. Let (ω_{ij}) be non-singular and let J be the Jacobi matrix of the transformation $(x) \to (y)$, then $A(x) = JA(y)J^T$, where $A(x) = (\omega_{ij}(x))$ and $A(y) = (\omega_{ij}(y))$, i.e. $\det A(x) = (\det J)^2 \cdot \det A(y) = (\det J)^2$

and, hence, $\det J = \sqrt{\det A(x)} = \sqrt{\det \omega_{ij}(x)}$. In the coordinates $y^1, \ldots, y^n$ the form Ω, at the point P, becomes $\Omega(y) = dy^1 \wedge \ldots \wedge dy^n$. Passing to $x^1, \ldots, x^n$, i.e. making the substitution $(x) \to (y)$, we obtain $\Omega(x) = \det\ J dx^1 \wedge \ldots \wedge dx^n$, so that $f(x) = \sqrt{\det \omega_{ij}(x)}$, which is what was required. Thus, the statement is proved for the non-singular matrix (ω_{ij}). If $\det \omega_{ij} = 0$, the matrix (ω_{ij}) does have zero eigenvalues, i.e. at the point P its canonical representation $\omega = \Sigma dx^i \wedge dx^{n+i}$ does not contain certain variables (corresponding to zero eigenvalues). Hence, considering the nth exterior degree of the form ω, we find $\Omega = 0$. The theorem is proved.

For $n = 2$ we have $\omega^{(2)} = \omega_{12} dx^1 \wedge dx^2 + \omega_{34} dx^3 \wedge dx^4$,

$$\Omega^{(4)} = \omega \wedge \omega = \frac{1}{2!}(2g_{12}g_{34}\, dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4)$$

$$= (g_{12}g_{34})\, dx^1 \wedge \ldots \wedge dx^4,$$

$$f^2(x) = \det(\omega_{ij}) = (g_{12}g_{34})^2, \quad f(x) = g_{12}g_{34} = \sqrt{\det(\omega_{ij})}.$$

5.3. CONNECTION AND COVARIANT DIFFERENTIATION

5.3.1. THE DEFINITION AND PROPERTIES OF AFFINE CONNECTION

Thus far, we were dealing only with algebraic operations over tensors, while differentiation was not touched upon. And this is no accident; now we shall try to explain the reasons for our caution.

The necessity of differentiation of a tensor field frequently arises in applied problems (see, for example, the definition of the divergence of a vector field). First of all it is ordinary differentiation of the components of a tensor in curvilinear coordinates. Let us consider, for simplicity, a covariant field $T_{i_1 \ldots i_k}$. Fix a curvilinear coordinate system $x^1, \ldots, x^n$ and construct a set of functions $P_{\alpha; i_1 \ldots i_k} = \frac{\partial}{\partial x^\alpha} T_{i_1 \ldots i_k}$, i.e. partial derivatives (in the usual sense) of the components of a tensor field. Perform this operation in each coordinate system. This means that in each system with the set of the components $T_{i'_1 \ldots i'_k}$ we associate a new set $P_{\alpha'; i'_1 \ldots i'_k} = \frac{\partial}{\partial x^{\alpha'}} T_{i'_1 \ldots i'_k}$. Do these sets form a tensor field P? In other words, are they transformed under coordinate substitution like tensors? Let us verify this fact by

direct calculation. The substitution $(x) \to (x')$ leads to the transformation

$$P_{\alpha';\, i'_1 \ldots i'_k}$$

$$= \frac{\partial T_{i'_1 \ldots i'_k}}{\partial x^{\alpha'}} = \frac{\partial}{\partial x^{\alpha'}} \left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} T_{i_1 \ldots i_k} \right)$$

$$= \frac{\partial}{\partial x^{\alpha}} (T_{i_1 \ldots i_k}) \frac{\partial x^{\alpha}}{\partial x^{\alpha'}} \frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}}$$

$$+ T_{i_1 \ldots i_k} \frac{\partial}{\partial x^{\alpha}} \left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i_k}} \right)$$

$$= P_{\alpha;\, i_1 \ldots i_k} \frac{\partial x^{\alpha}}{\partial x^{\alpha'}} \frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}}$$

$$+ T_{i_1 \ldots i_k} \frac{\partial}{\partial x^{\alpha'}} \left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} \right).$$

Thus,

$$P_{\alpha';\, i'_1 \ldots i'_k} = P_{\alpha;\, i_1 \ldots i_k} \frac{\partial x^{\alpha}}{\partial x^{\alpha'}} \frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} + S_{\alpha';\, i'_1 \ldots i'_k}(T, x, x'),$$

where

$$S_{\alpha';\, i'_1 \ldots i'_k} = T_{i_1 \ldots i_k} \frac{\partial}{\partial x^{\alpha'}} \left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} \right).$$

For an arbitrary substitution $(x) \to (x')$ the quantity $S_{\alpha';\, i'_1 \ldots i'_k}$ is, in general, non-zero, i.e. $P_{\alpha;\, i_1 \ldots i_k}$ are not transformed according to the tensor law. Hence, ordinary (non-covariant) differentiation $\frac{\partial}{\partial x^i}$ is not a tensor operation: it does not turn a tensor field into another tensor field. The expression for $S_{\alpha';\, i'_1 \ldots i'_k}$ also shows what assumptions are needed for the operation $\frac{\partial}{\partial x^i}$ to be a tensor one: $\frac{\partial}{\partial x^i}$ is a tensor operation only for a linear substitution $(x) \to (x')$; in this case $S_{\alpha';\, i'_1 \ldots i'_k} \equiv 0$. To bring further «discredit» (from the tensor standpoint) on the operation $\frac{\partial}{\partial x^i}$, we shall consider a vector field T^i referred to a curvilinear coordinate system (x) and calculate the divergence of this field by the formula $\operatorname{div} \mathbf{T} =$

$\sum_i \frac{\partial T^i}{\partial x^i}$ which is an analogue of the definition of divergence in a Euclidean coordinate system. Is this expression a scalar, i.e. is it invariant under an arbitrary substitution $(x) \to (x')$? Let us verify the equality $\operatorname{div} \mathbf{T}_{(x)} = \operatorname{div} \mathbf{T}_{(x')}$, where (x) and (x') are arbitrary coordinate systems. We have

$$\begin{aligned}\operatorname{div} \mathbf{T}_{(x')} &= \sum_i \frac{\partial T^{i'}}{\partial x^{i'}} = \frac{\partial}{\partial x^{i'}} \left(\frac{\partial x^{i'}}{\partial x^i} T^i \right) \\ &= \frac{\partial T^i}{\partial x^k} \frac{\partial x^k}{\partial x^{i'}} \frac{\partial x^{i'}}{\partial x^i} + T^i \frac{\partial x^k}{\partial x^{i'}} \frac{\partial^2 x^{i'}}{\partial x^k \partial x^i} \\ &= \sum_i \frac{\partial T^i}{\partial x^i} + R(\mathbf{T}, x, x') = \operatorname{div} \mathbf{T}_{(x)} + R(\mathbf{T}, x, x'),\end{aligned}$$

where

$$R(\mathrm{T}, x, x') = \mathrm{T}^i \frac{\partial x^k}{\partial x^{i'}} \cdot \frac{\partial^2 x^{i'}}{\partial x^k \partial x^i}.$$

For an arbitrary coordinate substitution $R(\mathbf{T}, x, x')$ does not vanish, i.e. the definition of divergence given above is not invariant. The explicit relation for $R(\mathbf{T}, x, x')$ shows that $R = 0$, for example, under linear substitutions. Thus, under non-linear substitutions $\frac{\partial}{\partial x^i}$ is not a tensor operation.

Our aim is to learn how to differentiate invariantly tensor fields in curvilinear coordinates. This is necessary if only because local coordinates on a smooth manifold are "almost always" curvilinear. Thus, it is required to find an operation (denoted by ∇, nabla) having the following properties (we assume for the time being that $M^n = \mathbf{R}^n$):

(1) in Cartesian coordinates in $\mathbf{R}^n$ the operation ∇ must coincide with ordinary differentiation $\left\{\frac{\partial}{\partial x^i}\right\}$,

(2) the operation ∇ must be a tensor one, i.e. if T is a tensor field on $\mathbf{R}^n$ (referred to curvilinear coordinates), ∇T is also a tensor field.

We begin with examples. Let us consider in $\mathbf{R}^n$ a vector field T^i; let (x) be a Cartesian coordinate system and (x') an arbitrary system. Let us write out conditions (1) and (2) imposed on ∇. In the system (x) we have $(\nabla T)^i_j = \frac{\partial T^i}{\partial x^j}$, and in the system (x') we have $(\nabla T)^{i'}_{j'} = \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^j}{\partial x^{j'}} (\nabla T)^i_j$. The problem is therefore to find

the explicit form of ∇ and calculate the components of $(\nabla T)^{i'}_{j'}$ in an arbitrary coordinate system. Straightforward calculation yields

$$(\nabla T)^{i'}_{j'} = \frac{\partial x^{i'}}{\partial x^l}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial}{\partial x^j}\left(\frac{\partial x^l}{\partial x^{k'}}T^{k'}\right)$$

$$= \frac{\partial x^{i'}}{\partial x^l}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^l}{\partial x^{k'}}\frac{\partial T^{k'}}{\partial x^{\alpha'}}\frac{\partial x^{\alpha'}}{\partial x^j}$$

$$+\frac{\partial x^{i'}}{\partial x^l}\frac{\partial x^j}{\partial x^{j'}}T^{k'}\frac{\partial}{\partial x^j}\left(\frac{\partial x^l}{\partial x^{k'}}\right)$$

$$= \delta^{i'}_{k'}\delta^{\alpha'}_{j'}\frac{\partial T^{k'}}{\partial x^{\alpha'}} + T^{k'}\frac{\partial x^{i'}}{\partial x^l}\frac{\partial^2 x^l}{\partial x^{j'}\partial x^{k'}},$$

$$(\nabla T)^{i'}_{j'} = \frac{\partial T^{i'}}{\partial x^{j'}} + T^{k'}\Gamma^{i'}_{j'k'}, \quad \Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^l}\cdot\frac{\partial^2 x^l}{\partial x^{j'}\partial x^{k'}}.$$

Thus, we have obtained the functions $\Gamma^{i'}_{j'k'}$ which measure the deviation of ∇ from ordinary (Euclidean) differentiation, i.e. we have learned how ∇ acts on a vector field.

Let us now consider a covector field T_i in $\mathbf{R}^n$. To find the explicit form of ∇ on T_i it is required, as before, to solve the system of equations

$$(\nabla T)_{ij} = \frac{\partial T_i}{\partial x_j}; \quad (\nabla T)_{i'j'} = \frac{\partial x^i}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}}(\nabla T)_{ij}.$$

We have

$$(\nabla T)_{i'j'} = \frac{\partial x^l}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial}{\partial x^j}\left(\frac{\partial x^{k'}}{\partial x^l}T_{k'}\right)$$

$$= \frac{\partial x^l}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^{k'}}{\partial x^l}\frac{\partial T_{k'}}{\partial x^{\alpha'}}\frac{\partial x^{\alpha'}}{\partial x^j} + \frac{\partial x^l}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}}T_{k'}\frac{\partial}{\partial x^j}\left(\frac{\partial x^{k'}}{\partial x^l}\right)$$

$$= \delta^{k'}_{i'}\delta^{\alpha'}_{j'}\frac{\partial T_{k'}}{\partial x^{\alpha'}} + T_{k'}\frac{\partial^2 x^{k'}}{\partial x^j\partial x^l}\cdot\frac{\partial x^l}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{j'}} = \frac{\partial T_{i'}}{\partial x^{j'}} + T_{k'}\widetilde{\Gamma}^{k'}_{i'j'},$$

where $\widetilde{\Gamma}^{k'}_{i'j'} = \frac{\partial^2 x^{k'}}{\partial x^j\partial x^l}\cdot\frac{\partial x^l}{\partial x^{i'}}\cdot\frac{\partial x^j}{\partial x^{j'}}$. Thus, $(\nabla T)_{i'j'} = \frac{\partial T_{i'}}{\partial x^{j'}} + T_{k'}\widetilde{\Gamma}^{k'}_{i'j'}$,

where $\widetilde{\Gamma}^{k'}_{i'j'}$ are the functions which measure the deviation of ∇ from ordinary differentiation on a covector field.

Lemma 1. *The equality* $\widetilde{\Gamma}^{k'}_{i'j'} = -\Gamma^{k'}_{i'j'}$ *is valid.*

Proof. The identity $\frac{\partial x^{i'}}{\partial x^{i''}} \cdot \frac{\partial x^{i''}}{\partial x^{k'}} = \delta^{i'}_{k'}$ is obvious. Differentiation with respect to $x^{p'}$ yields

$$\frac{\partial^2 x^{i''}}{\partial x^{p'}\, \partial x^{k'}} \cdot \frac{\partial x^{i'}}{\partial x^{i''}} + \frac{\partial x^{i''}}{\partial x^{k'}} \cdot \frac{\partial^2 x^{i'}}{\partial x^{p''} \partial x^{i''}} \cdot \frac{\partial x^{p''}}{\partial x^{p'}} = 0,$$

i.e. $\Gamma^{i'}_{p'k'} + \widetilde{\Gamma}^{i'}_{p'k'} = 0$. The lemma is proved.

Thus, the action of ∇ on vector and covector fields (in curvilinear coordinates in $\mathbf{R}^n$) is of the form

$$(\nabla T)^{i'}_{j'} = \frac{\partial T^{i'}}{\partial x^{j'}} + T^{k'}\Gamma^{i'}_{k'j'},$$

$$(\nabla T)_{i'j'} = \frac{\partial T_{i'}}{\partial x^{j'}} - T_{k'}\Gamma^{k'}_{i'j'}.$$

Let us now consider the action of ∇ on operator fields, i.e. on fields of type (1, 1). We have

$$(\nabla T)^{i}_{jk} = \frac{\partial}{\partial x^k} (T^i_j),$$

$$\begin{aligned}
(\nabla T)^{i'}_{j'k'} &= \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \frac{\partial}{\partial x^k} \left(\frac{\partial x^i}{\partial x^{\alpha'}} \frac{\partial x^{p'}}{\partial x^j} T^{\alpha'}_{p'} \right) \\
&= \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \frac{\partial x^i}{\partial x^{\alpha'}} \frac{\partial x^{p'}}{\partial x^j} \frac{\partial T^{\alpha'}_{p'}}{\partial x^{q'}} \frac{\partial x^{q'}}{\partial x^k} \\
&\quad + T^{\alpha'}_{p'} \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \cdot \frac{\partial^2 x^i}{\partial x^{q'}\, \partial x^{\alpha'}} \frac{\partial x^{q'}}{\partial x^k} \frac{\partial x^{p'}}{\partial x^j} \\
&\quad + T^{\alpha'}_{p'} \frac{\partial x^{i'}}{\partial x^i} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \frac{\partial x^i}{\partial x^{\alpha'}} \frac{\partial^2 x^{p'}}{\partial x^k\, \partial x^j} \\
&= \frac{\partial}{\partial x^{k'}} (T^{i'}_{j'}) + T^{p'}_{j'}\Gamma^{i'}_{p'k'} - T^{i'}_{p'}\Gamma^{p'}_{j'k'}.
\end{aligned}$$

Thus, we have found how ∇ acts on T^i_j.

Theorem 1. *Let $M^n = \mathbf{R}^n$, let (x) be Cartesian coordinates, and let (x') be arbitrary curvilinear coordinates. Then there exists in $\mathbf{R}^n$ a tensor operation ∇ defined on an arbitrary tensor field $T^{i_1 \ldots i_k}_{j_1 \ldots j_p}$ by the formula*

$$(\nabla T)^{i'_1 \ldots i'_k}_{j'_1 \ldots j'_p;\alpha'} = \frac{\partial}{\partial x^{\alpha'}} \left(T^{i'_1 \ldots i'_k}_{j'_1 \ldots j'_p} \right) + \sum_{s=1}^{k} T^{i'_1 \ldots i'_s = q' \ldots i'_k}_{j'_1 \ldots j'_p} \Gamma^{i'_s}_{q'\alpha'}$$
$$- \sum_{s=1}^{p} T^{i'_1 \ldots i'_k}_{j'_1 \ldots j'_s = q' \ldots j'_p} \Gamma^{q'}_{j_s\alpha'},$$

where the functions $\Gamma^{i'}_{j'q'}$ are transformed under the substitution $(x') \to (x'')$ as

$$\Gamma^{i''}_{j''k''} = \frac{\partial x^{i''}}{\partial x^{i'}} \frac{\partial x^{j'}}{\partial x^{j''}} \frac{\partial x^{k'}}{\partial x^{k''}} \Gamma^{i'}_{j'k'} + \frac{\partial x^{i''}}{\partial x^{i'}} \frac{\partial^2 x^{i'}}{\partial x^{j''} \partial x^{k''}}.$$

Proof. The explicit expression for ∇ on the field $T^{i_1 \ldots i_k}_{j_1 \ldots j_p}$ is obtained by making the above calculations for vector, covector, and operator fields as many times as the rank of $T^{i_1 \ldots i_k}_{j_1 \ldots j_p}$. As a useful exercise, we recommend the reader perform all the calculations. Let us examine the transformation law for $\Gamma^{i'}_{j'k'}$. We have

$$\nabla_{k'} T^{i'} = \frac{\partial T^{i'}}{\partial x^{k'}} + T^{p'} \Gamma^{i'}_{p'k'},$$

$$\begin{aligned} \nabla_{k''} T^{i''} &= \frac{\partial T^{i''}}{\partial x^{k''}} + T^{p''} \Gamma^{i''}_{p''k''} \\ &= \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial}{\partial x^{k'}} \left(\frac{\partial x^{i''}}{\partial x^{i'}} T^{i'} \right) + \frac{\partial x^{p''}}{\partial x^{p'}} T^{p'} \Gamma^{i''}_{p''k''} \\ &= \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial x^{i''}}{\partial x^{i'}} \cdot \frac{\partial T^{i'}}{\partial x^{k'}} + T^{i'} \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial^2 x^{i''}}{\partial x^{k'} \partial x^{i''}} + T^{p'} \frac{\partial x^{p''}}{\partial x^{p'}} \Gamma^{i''}_{p''k''} \end{aligned}$$

$$\begin{aligned} \nabla_{k''} T^{i''} &= \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial x^{i''}}{\partial x^{i'}} \cdot \nabla_{k'} T^{i'} \\ &= \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial x^{i''}}{\partial x^{i'}} \left(\frac{\partial T^{i'}}{\partial x^{k'}} + T^{p'} \Gamma^{i'}_{p'k'} \right). \end{aligned}$$

Comparison of these two equalities shows that

$$T^{p'} \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial x^{i''}}{\partial x^{i'}} \Gamma^{i'}_{p'k'} = T^{p'} \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial^2 x^{i''}}{\partial x^{k'} \partial x^{p'}} + T^{p'} \frac{\partial x^{p''}}{\partial x^{p'}} \Gamma^{i''}_{p''k''}.$$

Since this identity must be satisfied for any field T^i, we obtain

$$\Gamma^{i''}_{p''k''} = \Gamma^{i'}_{p'k'} \frac{\partial x^{p'}}{\partial x^{p''}} \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial x^{i''}}{\partial x^{i}} - \frac{\partial x^{p'}}{\partial x^{p''}} \frac{\partial x^{k'}}{\partial x^{k''}} \frac{\partial^2 x^{i''}}{\partial x^{k'} \partial x^{p'}}.$$

According to Lemma 1,

$$\frac{\partial^2 x^{i''}}{\partial x^{k'} \partial x^{p'}} \cdot \frac{\partial x^{p'}}{\partial x^{p''}} \cdot \frac{\partial x^{k'}}{\partial x^{k''}} = \frac{\partial^2 x^{k'}}{\partial x^{p''} \partial x^{k''}} \cdot \frac{\partial x^{i''}}{\partial x^{k'}} = \frac{\partial^2 x^{i'}}{\partial x^{p''} \partial x^{k''}} \cdot \frac{\partial x^{i''}}{\partial x^{i'}}.$$

The theorem is proved.

We have proved the theorem on the existence of "tensor differentiation", but only for the case $M^n = \mathbf{R}^n$, i.e. when M^n is provided

with Cartesian coordinates. This has enabled us to calculate explicitly the functions $\Gamma^{i'}_{j'k'}$ which measure the deviation of ∇ from ordinary Euclidean differentiation $\frac{\partial}{\partial x^{i'}}$, namely,

$$\Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{i}} \frac{\partial^2 x^i}{\partial x^{j'} \partial x^{k'}} = -\frac{\partial^2 x^{k'}}{\partial x^j \partial x^l} \cdot \frac{\partial x^l}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} .$$

We tacitly assumed that in $\mathbf{R}^n$ there exist "preferred" Cartesian coordinates in which the operation ∇ coincides with ordinary differentiation. Let us now turn to an arbitrary smooth manifold and define the operation ∇ axiomatically, relying upon the properties of ∇ in $\mathbf{R}^n$ demonstrated above.

Definition. *Covariant differentiation* ∇ is said to be defined on a smooth manifold M^n if for each smooth atlas there is given in each chart a set of smooth functions Γ^i_{jk} transformed under coordinate substitution according to the law

$$\Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{i}} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \Gamma^i_{jk} + \frac{\partial x^{i'}}{\partial x^{l}} \cdot \frac{\partial^2 x^l}{\partial x^{j'} \partial x^{k'}} .$$

Then ∇ is given by the formula

$$(\nabla T)^{i_1 \ldots i_k}_{j_1 \ldots j_p;\alpha} = \frac{\partial}{\partial x^\alpha} \left(T^{i_1 \ldots i_k}_{j_1 \ldots j_p} \right) + \sum_{s=1}^{k} T^{i_1 \ldots; i_s = q; \ldots i_k}_{j_1 \ldots j_p} \Gamma^{i_s}_{q\alpha}$$

$$- \sum_{s=1}^{p} T^{i_1 \ldots i_k}_{j_1 \ldots; j_s = q; \ldots j_p} \Gamma^{q}_{j_s \alpha} .$$

The existence of manifolds M^n on which the operation ∇ is valid has already been proved. We can therefore set $M^n = \mathbf{R}^n$; then Γ^i_{jk} are given by the formulas derived above. Essentially, the set of Γ^i_{jk} does not form a tensor. These functions are called *Christoffel symbols*. They are transformed by the tensor law if the substitution $(x) \to (x')$ is linear. In the definition of ∇ the system (x) (not primed) is no longer a Cartesian one because we now deal with an arbitrary manifold M^n. The Christoffel symbols (or ∇) are sometimes said to define an *affine connection* on M^n.

Definition. The *torsion tensor of the affine connection* Γ^i_{jk} is a tensor defined in each coordinate system by the relation $\Omega^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj}$.

Lemma 2. *The set of functions* Ω^i_{jk} *does form a tensor.*

Proof. Under coordinate substitution, Γ^i_{jk} are transformed as

$$\Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{l}} \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \Gamma^l_{jk} + \frac{\partial x^{i'}}{\partial x^{l}} \frac{\partial^2 x^l}{\partial x^{j'} \partial x^{k'}} .$$

Alternating $\Gamma^{i'}_{j'k'}$ with respect to the lower indices and taking into account that the "non-tensor term" is symmetric in j' and k' we obtain

$$\Omega^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{i}} \frac{\partial x^{j}}{\partial x^{j'}} \cdot \frac{\partial x^{k}}{\partial x^{k'}} \Omega^{i}_{jk}.$$

The lemma is proved.

Definition. The connection (or covariant differentiation) ∇, is called *symmetric* if the torsion tensor is zero.

The connection introduced in $\mathbf{R}^n$ is symmetric; this follows from the expressions for ∇ in terms of partial derivatives. On an arbitrary M^n, the Christoffel symbols need not necessarily have the form of second-order partial derivatives of the coordinates; such a form arises if only Euclidean coordinates exist locally on M^n. In a fixed coordinate system, ∇ is represented as a combination of operations ∇_k, covariant derivatives with respect to individual coordinates (analogues of partial derivatives $\frac{\partial}{\partial x^k}\Big)$.

Proposition 1. *Covariant differentiation (connection) ∇ satisfies the following relations:*

(1) *the operation $\nabla = \{\nabla_k\}$ is linear,*

(2) *for an arbitrary tensor field $T^{(i)}_{(j)}$ the set of functions $\nabla_k T^{(i)}_{(j)} = (\nabla T)^{(i)}_{k,(j)}$ forms a tensor field,*

(3) *if the tensor field is scalar (i.e. a smooth function f on M^n), then $\nabla f = \{\nabla_k f\} = \left\{\frac{\partial f}{\partial x^k}\right\} = \operatorname{grad} f$,*

(4) *on a vector field T^i the operation ∇ is of the forms,*

$$\nabla_k T^i = \frac{\partial T^i}{\partial x^k} + T^\alpha \Gamma^i_{\alpha k},$$

and on a covector field T_i the operation ∇ is given by

$$\nabla_k T_i = \frac{\partial T_i}{\partial x^k} - T_\alpha \Gamma^\alpha_{ik},$$

(5) *the operation ∇ satisfies the Leibniz formula*

$$\nabla_k \{T^{(i)}_{(j)} \cdot P^{(\alpha)}_{(\beta)}\} = (\nabla_k T^{(i)}_{(j)}) P^{(\alpha)}_{(\beta)} + T^{(i)}_{(j)} \cdot (\nabla_k P^{(\alpha)}_{(\beta)}),$$

where $T^{(i)}_{(j)}$ and $P^{(\alpha)}_{(\beta)}$ are arbitrary tensor fields.

Proof. Properties (1)-(4) immediately follow from the definition of ∇. It remains to prove (5). Let us consider the simplest case: one of the fields, $T^{(i)}_{(j)}$ or $P^{(\alpha)}_{(\beta)}$, is scalar. Then (5) follows from the

Leibniz formula for scalar functions. Let now T and P be vector fields T^i, P^j. We have

$$\begin{aligned}\nabla_k (T^i \cdot P^j) &= \frac{\partial}{\partial x^k} (T^i \cdot P^j) + T^\alpha P^j \Gamma^i_{\alpha k} + T^i P^\alpha \Gamma^j_{\alpha k} \\ &= \left(\frac{\partial}{\partial x^k} T^i \right) P^j + T^i \frac{\partial}{\partial x^k} (P^j) + T^\alpha P^j \Gamma^i_{\alpha k} \\ &= \left(\frac{\partial T^i}{\partial x^k} + T^\alpha \Gamma^i_{\alpha k} \right) P^j + T^i \left(\frac{\partial P^j}{\partial x^k} + P^\alpha \Gamma^j_{\alpha k} \right) \\ &= (\nabla_k T^i) P^j + T^i (\nabla_k P^j).\end{aligned}$$

For arbitrary fields P and T property (5) is proved by repeating the above reasoning (after the substitution of multi-indices (i), (j) for i, j) and using the definition of ∇. It is left for the reader as a useful exercise to perform all calculations. The proposition is proved.

Theorem 2. *Given on M^n the operation $\nabla = \{\nabla_k\}$ satisfying* (1)-(5) (*see Proposition 1*). *Then for an arbitrary tensor field $T^{(i)}_{(j)}$ the following identity is valid*:

$$\nabla_k T^{(i)}_{(j)} = \nabla_k T^{i_1 \ldots i_\alpha}_{j_1 \ldots j_p} = \frac{\partial}{\partial x^k} \left(T^{i_1 \ldots i_\alpha}_{j_1 \ldots j_p} \right) + \sum_{q=1}^{\alpha} T^{i_1 \ldots ; i_q = r; \ldots i_\alpha}_{j_1 \ldots j_p} \cdot \Gamma^{i_q}_{rk}$$

$$- \sum_{s=1}^{p} T^{i_1 \ldots i_\alpha}_{j_1 \ldots ; j_s = \omega; \ldots j_p} \Gamma^{\omega}_{j_s k},$$

i.e. ∇ is covariant differentiation in the sense of our definition (see above). The algebraic properties (1)-(5) *uniquely define the operation ∇, i.e. it can be introduced axiomatically with the aid of properties* (1)-(5).

Proof. On scalar functions and on tensor fields of rank 1 the operation ∇_k is given by (1)-(4). It remains to define ∇_k on an arbitrary tensor field. We now prove an auxiliary *lemma*: any tensor field can be decomposed into a linear combination (with smooth coefficients) of the products of first-rank fields. Indeed, any tensor field is a multilinear mapping

$$T = a^{i_1 \ldots i_k}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \mathbf{e}^{j_1} \otimes \ldots \otimes \mathbf{e}^{j_q} \colon (T_*)^q \times (T^*)^k \to \mathbf{R},$$

where $\mathbf{e}_{i_\alpha}$, $\mathbf{e}^{i_\alpha}$ are fields of rank 1, and $a^{i_1 \ldots i_k}_{j_1 \ldots j_q}$ are smooth functions. This definition is invariant, so that the decomposition refers to an arbitrary fixed coordinate system (in another system the decomposition will change). Let us choose a coordinate system $x^1, \ldots, x^n$ and let $\{T^{\alpha}_{i_s}\}$ (i_s is fixed) be the set of the components of the tensor

$\mathbf{e}_{i_s}$, and, correspondingly, let $T^{j_s}_{\beta}$ (j_s is fixed) be the set of the components of the tensor $\mathbf{e}^{j_s}$. Expressing T in this system, we obtain

$$T = a^{i_1 \dots i_k}_{j_1 \dots j_q} T^{\alpha_1}_{i_1} \dots T^{\alpha_k}_{i_k} T^{j_1}_{\beta_1} \dots T^{j_q}_{\beta_q} \left(\frac{\partial}{\partial x^{\alpha_1}}\right) \otimes \dots \otimes \left(\frac{\partial}{\partial x^{\alpha_k}}\right) \otimes \left(\frac{\partial}{\partial x}\right)^{\beta_1} \otimes \dots \otimes \left(\frac{\partial}{\partial x}\right)^{\beta_q};$$

i.e.

$$T = T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_q} \left(\frac{\partial}{\partial x^{\alpha_1}}\right) \otimes \dots \otimes \left(\frac{\partial}{\partial x^{\alpha_k}}\right) \otimes \left(\frac{\partial}{\partial x}\right)^{\beta_1} \otimes \dots \otimes \left(\frac{\partial}{\partial x}\right)^{\beta_q},$$

where

$$T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_q} = a^{i_1 \dots i_k}_{j_1 \dots j_q} T^{\alpha_1}_{i_1} \dots T^{\alpha_k}_{i_k} T^{j_1}_{\beta_1} \dots T^{j_q}_{\beta_q}.$$

Thus, any multilinear mapping is completely defined by the coefficients of decomposition relative to the basis. Let us now return to the formula for the operator ∇ on $T^{(i)}_{(j)}$. Consider, for simplicity, a field of rank 2 (for an arbitrary field calculations are analogous). Since ∇ is linear, it is sufficient to verify formula (5) for the monomials $T_{ij} = \alpha(x) T_i P_j$, where $\alpha(x)$ is a smooth function. Calculate $\nabla_k (T_{ij})$ under the assumption that $\nabla_k (T_i)$ are already known (see (4)). By the Leibniz formula (5) we have

$$\begin{aligned}
\nabla_k (T_{ij}) &= \nabla_k (\alpha(x) T_i P_j) \\
&= \frac{\partial \alpha}{\partial x^k} T_i P_j + \alpha \left(\frac{\partial T_i}{\partial x^k} - T_p \Gamma^p_{ik}\right) P_j + \alpha T_i \left(\frac{\partial P_j}{\partial x^k} - P_q \Gamma^q_{jk}\right) \\
&= \frac{\partial}{\partial x^k} (\alpha T_i P_j) - \alpha \Gamma^p_{ik} T_p P_j - \alpha \Gamma^p_{jk} T_i P_q \\
&= \frac{\partial}{\partial x^k} (T_{ij}) - \Gamma^q_{ik} T_{pj} - \Gamma^q_{jk} T_{iq},
\end{aligned}$$

which is what was required. The theorem is proved.

Definition. Let ∇ be an affine connection on M^n. Local coordinates $x^1, \dots, x^n$ are called *Euclidean* for ∇ if in these coordinates $\Gamma^i_{jk}(x) = 0$.

Such coordinates may not exist if, for example, ∇ is non-symmetric. Indeed, in this case Ω^i_{jk} is not zero. If there existed the coordinates in which $\Gamma^i_{jk}(x) \equiv 0$, the torsion tensor Ω^i_{jk} would vanish in this coordinate system and therefore it would be identically zero in any coordinate system (because Ω^i_{jk} is a tensor).

Essentially, the concept of covariant differentiation (affine connection) does not rely upon the concept of a Riemannian metric; the only fact we have used is the existence of a smooth structure on M^n. The Riemannian metric and connection are, therefore, two distinct structures on M^n. In particular, Euclidean coordinates for the affine connection and Euclidean coordinates for the metric g_{ij} (i.e. the coordinates in which $g_{ij} = \delta_{ij}$) are distinct concepts. Thus, the Riemannian metric and covariant differentiation do not, in general, define each other.

Let a metric g_{ij} be given on M^n. Then, among all affine symmetric connections we can choose one (and only one!) connection which is "consistent" with this metric and is completely defined by this metric. This is an important class of connections called Riemannian connections.

5.3.2. RIEMANNIAN CONNECTIONS

Let g_{ij} be a metric and $\nabla = \{\Gamma^i_{jk}\}$ an affine connection on M^n.

Definition. The affine symmetric connection $\nabla = \{\Gamma^i_{jk}\}$ is called *compatible with the metric* g_{ij} (or the *Riemannian connection*) if $\nabla(g_{ij}) \equiv 0$.

We know that ∇ is a tensor, and this means that if the identity $\{\nabla_k(g_{ij}) \equiv 0\}$ holds in one coordinate system, it will also hold in all other systems. Thus, the tensor g_{ij} is "constant" with respect to the Riemannian connection in the sense that the covariant derivative of this tensor is zero. It follows (for any tensor field) that $T^{(p)}_{(q)}$: $\nabla_k (g_{ij}T^{(p)}_{(q)}) \equiv g_{ij}\nabla_k (T^{(p)}_{(q)})$, see the Leibniz formula (Sec. 5.3.1, Proposition 1, relation (5)). In particular, ∇ commutes with raising and lowering the indices. We now prove that Riemannian connection does exist and is unique.

Theorem 3. *Let g_{ij} be a metric on M^n. Then there exists a unique symmetric affine connection compatible with g_{ij} and such that*

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\alpha} \left(\frac{\partial g_{k\alpha}}{\partial x^j} + \frac{\partial g_{j\alpha}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^\alpha} \right).$$

Proof. Suppose the existence is proved. Let us demonstrate that the connection is unique. By definition, $\nabla_k (g_{ij}) = 0$. Since ∇ is defined in terms of Γ^i_{jk}, it is sufficient to prove that this system uniquely gives Γ^i_{jk} as functions of g_{ij} and $\frac{\partial g_{ij}}{\partial x^\alpha}$. We have $\nabla_k g_{ij} = 0$, $\frac{\partial g_{ij}}{\partial x^k} = g_{\alpha j}\Gamma^\alpha_{ik} + g_{\alpha i}\Gamma^\alpha_{jk}$. Circular permu-

tation of indices yields

$$\begin{array}{rl|l} + & \dfrac{\partial g_{ij}}{\partial x^k} = \Gamma_{j,ik} + \Gamma_{i,jk} & (ijk) \\ + & \dfrac{\partial g_{ki}}{\partial x^j} = \Gamma_{i,kj} + \Gamma_{k,ij} & (kij) \\ - & \dfrac{\partial g_{jk}}{\partial x^i} = \Gamma_{k,ji} + \Gamma_{j,ki} & (jki) \end{array}$$

where $\Gamma_{i,\,kj} = g_{i\alpha}\Gamma^{\alpha}_{kj}$. Adding the first two identities and subtracting the third one from the sum, we obtain

$$\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i} = 2\Gamma_{i,jk} = 2g_{i\alpha}\Gamma^{\alpha}_{jk}.$$

(recall that $\Gamma^i_{kj} = \Gamma^i_{jk}$). Hence,

$$\Gamma^{\alpha}_{jk} = \frac{1}{2} g^{i\alpha}\left(\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i}\right).$$

We have used the rule for solving the system $g_{\alpha\beta} = T^{\alpha} = Q_{\beta}$. Since $g_{\alpha\beta}$ is a non-singular tensor, there exists the inverse tensor $g^{\alpha\beta}$ such that $g_{\alpha\beta}g^{\beta\gamma} = \delta^{\gamma}_{\alpha}$, whence

$$g^{\beta\gamma}g_{\alpha\beta}T^{\alpha} = g^{\beta\gamma}Q_{\beta},\ \delta^{\gamma}_{\alpha}T^{\alpha} = g^{\beta\gamma}Q_{\beta},\ T^{\alpha} = g^{\beta\alpha}Q_{\beta}.$$

We have obtained the formulas which express Γ^{α}_{jk} in terms of g_{ij} and its derivatives. This means that if the initial system has a solution, it is unique. To prove the existence of the connection, it is sufficient to calculate Γ^i_{jk} by the formulas derived above. Reversing the manipulations, we obtain $\nabla_k(g_{ij}) \equiv 0$. The theorem is proved.

Remark. Let, on M^n, there exist a coordinate system $x^1, \ldots, x^n$ in which $g_{ij} = \delta_{ij}$. The existence of such a system is equivalent to the existence of the system in which (g_{ij}) is a constant matrix (independent of the point in some neighbourhood). Then under a linear transformation, this matrix can be reduced (simultaneously at all points of the neighbourhood) to the form δ_{ij}. Thus, a coordinate system may be called Euclidean if in this system the matrix of the metric tensor is constant. With this remark in mind, we turn back to an analysis of the Riemannian connection. A local coordinate system is Euclidean, from the connection standpoint, if and only if in this system $\{\Gamma^i_{jk} \equiv 0\}$ (in some neighbourhood). **Statement:** A coordinate system is Euclidean, from the standpoint of the Riemannian connection compatible with g_{ij} if and only if this system is Euclidean from the standpoint of g_{ij} (i.e. g_{ij} is locally constant). Indeed, if g_{ij} is locally constant, then $\Gamma^i_{jk} \equiv 0$, by Theorem 3, i.e.

this system is Euclidean for the Riemannian connection ∇ as well. Conversely, if $\Gamma^i_{jk}=0$, then, by Theorem 3, $\frac{\partial g_{ij}}{\partial x^k}=g_{\alpha j}\Gamma^\alpha_{ik}+g_{\alpha i}\Gamma^\alpha_{jk}=0$, i.e. g_{ij} is constant in this system, which is what was required.

Let us consider a smooth hypersurface $V^{n-1}\subset \mathbf{R}^n$ defined by the graph $x^n=f(x^1, \ldots, x^{n-1})$ and let $P_0\in V^{n-1}$ be an arbitrary point. Let also $T_{P_0}(V^{n-1})$ be parallel to the plane $\mathbf{R}^{n-1}(x^1, \ldots, x^{n-1})$, i.e. $\frac{\partial f(P_0)}{\partial x^i}=0$, $1\leqslant i\leqslant n-1$. We have already calculated g_{ij} at P_0 relative to this special coordinate system. We have $g_{ij}=\delta_{ij}+f_{x^i}+f_{x^j}$, whence $\left.\frac{\partial g_{ij}}{\partial x^k}\right|_{P_0}=0$, i.e. $\Gamma^i_{jk}(P_0)=0$, where ∇ is the Riemannian connection compatible with g_{ij}. Generally, Γ^i_{jk} are equal to zero only at the point P_0; Γ^i_{jk} need not necessarily be zero in a neighbourhood of P_0.

We have already noted that an attempt at defining the divergence of a vector field by the "Euclidean formula" $\sum_{i=1}^{n}\frac{\partial}{\partial x^i}\ (T^i)$ generally fails because this expression is not a scalar. We now demonstrate that the existence of a Riemannian connection permits a correct definition of the flow divergence. Note that the concept of divergence (a change in an infinitesimal volume) implies the existence of a metric on M^n, otherwise the concept of volume is devoid of sense.

Let $\mathbf{T}$ be a vector field on M^n provided with Riemannian connection. Set $\operatorname{div}\mathbf{T}=\nabla_iT^i$. It follows from the properties of ∇ that $\operatorname{div}\mathbf{T}$ is a scalar function on M^n. Let us find the explicit form of $\operatorname{div}\mathbf{T}$ in terms of g_{ij} and the field components.

Statement 1. *The following relation is valid:*

$$\operatorname{div}\mathbf{T}=\frac{\partial T^i}{\partial x^i}+T^\alpha\frac{\partial}{\partial x^\alpha}(\ln\sqrt{g}),$$

where g is the determinant of the matrix (g_{ij}).

Proof. According to Theorem 3, we have

$$\begin{aligned}\operatorname{div}\mathbf{T}=\nabla_iT^i&=\frac{\partial T^i}{\partial x^i}+T^\alpha\Gamma^i_{\alpha i}\\ &=\frac{\partial T^i}{\partial x^i}+T^\alpha\cdot\frac{1}{2}g^{ip}\left(\frac{\partial g_{p\alpha}}{\partial x^i}+\frac{\partial g_{pi}}{\partial x^\alpha}-\frac{\partial g_{\alpha i}}{\partial x^p}\right),\end{aligned}$$

because

$$\Gamma^i_{\alpha k}=\frac{1}{8}g^{ip}\left(\frac{\partial g_{p\alpha}}{\partial x^k}+\frac{\partial g_{pk}}{\partial x^\alpha}-\frac{\partial g_{\alpha k}}{\partial x^p}\right).$$

Furthermore,

$$\operatorname{div} \mathbf{T} = \frac{\partial T^i}{\partial x^i} + \frac{1}{2} T^\alpha \left(g^{ip} \frac{\partial g_{p\alpha}}{\partial x^i} + g^{ip} \frac{\partial g_{pi}}{\partial x^\alpha} - g^{ip} \frac{\partial g_{\alpha i}}{\partial x^p} \right)$$
$$= \frac{\partial T^i}{\partial x^i} + \frac{1}{2} T^\alpha g^{ip} \frac{\partial g_{ip}}{\partial x^\alpha},$$

since

$$g^{ip} \frac{\partial g_{p\alpha}}{\partial x^i} - g^{ip} \frac{\partial g_{\alpha i}}{\partial x^p} = 0.$$

It is required to find $g^{ip} \frac{\partial g_{ip}}{\partial x^\alpha}$. We assert that

$$g^{ip} \frac{\partial g_{ip}}{\partial x^\alpha} = 2 \frac{\partial}{\partial x^\alpha} (\ln \sqrt{g}).$$

Recall that Δ_{ip} is the minor (with sign) complementary to g_{ip} in the matrix (g_{ij}). This implies that it is sufficient to verify the relation

$$\frac{1}{g} \sum_{(i,p)} \Delta_{ip} \frac{\partial g_{ip}}{\partial x^\alpha} = \frac{1}{\sqrt{g}} \frac{1}{\sqrt{g}} \frac{\partial g}{\partial x^\alpha},$$

i.e. to prove that

$$\sum_{(i,p)} \Delta_{ip} \frac{\partial g_{ip}}{\partial x^\alpha} = \frac{\partial g}{\partial x^\alpha}.$$

The determinant $g = \det g_{ij}$ is the sum of homogeneous monomials of degree n, $n = \dim M^n$. Fix g_{ip} in g and collect in the sum g all the terms containing g_{ip} (as a factor). Then $g = \ldots + g_{ip}R_{ip} + \ldots$, where R_{ip} is a polynomial of degree $n-1$ which does not include g_{ip}. The pair of indices (i, p) may be arbitrary, but in order to find the coefficient of another element $g_{\alpha\beta}$, we should "scatter" the preceding sum and arrange the determinant g as a new sum, picking up the terms with $g_{\alpha\beta}$: $g = \ldots + g_{\alpha\beta} R_{\alpha\beta} + \ldots$. Let the pair (i, p) be fixed; then $R_{ip} = \Delta_{ip}$. Indeed, consider the standard expansion of g by a column (or a row) $g = \sum_{\alpha=1}^{n} g_{i\alpha}\Delta_{i\alpha}$ (this is true for any fixed i). Since g_{ip} appears only in the term $g_{ip}\Delta_{ip}$ (in the sum $\sum_{\alpha=1}^{n} g_{i\alpha}\Delta_{i\alpha}$), $R_{ip} = \Delta_{ip}$. Calculating $\frac{\partial g}{\partial x}$, we see that $\frac{\partial g_{ip}}{\partial x^\alpha}$ appears in the final relation for g with the factor Δ_{ip}, while the remaining terms of this sum,

i.e. $(A_{ip}) = g - g_{ip}\Delta_{ip}$, do not contain the function g_{ip}. Hence, $\frac{\partial g}{\partial x^\alpha} = \sum_{(i,p)} \frac{\partial g_{ip}}{\partial x^\alpha}\Delta_{ip}$, which is what was required. Thus,

$$\operatorname{div}\mathbf{T} = \frac{\partial T^i}{\partial x^i} + T^\alpha \frac{\partial}{\partial x^\alpha}(\ln\sqrt{g}).$$

5.4. PARALLEL DISPLACEMENT. GEODESICS

5.4.1. PRELIMINARIES

Let us consider a smooth manifold (not necessarily a Riemannian one). In many particular problems there is a need to compare vectors applied at distinct points; for example, it is required to compare two tangent vectors lying in distinct tangent spaces. In the case of an arbitrary M^n such a comparison is rather difficult because T_xM^n and T_yM^n are distinct and may be identified in many different ways. In certain specific cases, where, for example, $M^n = \mathbf{R}^n$, we can use the operation of parallel displacement or translation which permits the comparison of vectors applied at distinct points. Formally, this procedure (for $M^n = \mathbf{R}^n$) can be defined as follows. Let us consider two points P and Q and let $\mathbf{a}$ be a vector at the point P. Consider also a smooth curve $\gamma(t)$ through these points, $\gamma(0) = P$ and $\gamma(1) = Q$, and translate the vector $\mathbf{a}$ along $\gamma(t)$ in such a manner that the vector remains parallel to itself. This operation generates along γ a vector field $\mathbf{a}(t)$ whose components are constant (with respect to t) and equal to the components of $\mathbf{a}$ at the initial moment. In particular, the derivatives of the components of the field $\mathbf{a}(t)$ with respect to t are zero. The vector $\mathbf{a}(1)$ applied at the point Q does not depend on the curve γ along which $\mathbf{a}$ is translated; we may therefore say that the parallel displacement in $\mathbf{R}^n$ does not depend on the path (Fig. 5.11).

For an arbitrary M^n, however, this simple scheme does not work. This is primarily because M^n cannot be covered with one chart, i.e. a unique coordinate system, common for all points, cannot be defined on M^n. Suppose first that M^2 is smoothly embedded in $\mathbf{R}^3$ and let $P, Q \in M^2$ be a pair of sufficiently close points on M^2. Let $\mathbf{a} \in T_PM^2$ and let γ be a curve from P to Q. We may suggest the following rule of parallel displacement on M^2: consider $\mathbf{a}$ as a vector in $\mathbf{R}^3$ and effect along γ an ordinary "three-dimensional" translation. In this case $\mathbf{a}$ turns into $\mathbf{b}$ at the point Q, but $\mathbf{b}$ need not belong to $T_Q M^2$. This drawback can be eliminated by orthogonally projecting $\mathbf{b}$ onto $T_Q M^2$ and defining the projection $\pi\mathbf{b}$ as the parallel displacement of $\mathbf{a}$ from P to Q along γ (see Fig. 5.12). The result of this operation does not depend on the translation path, but the operation has a drawback; it is defined only in a small neighbourhood of P. If

we translate **a** "by a larger distance", the vector $\pi\mathbf{b}$, which was called "parallel to **a**", may be a zero one. This occurs, for example, on S^2 (see Fig. 5.13) if the translation path is a quarter of the meridian PQ. There are many arguments, according to which a parallel

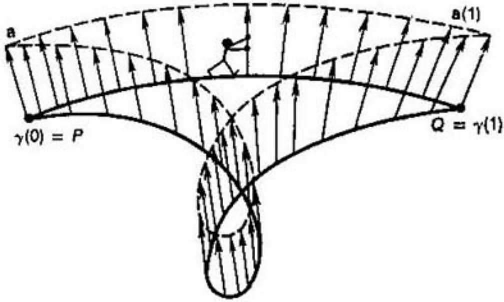

Figure 5.11

displacement leading to a zero vector should be rejected. The procedure may, of course, be more accurate: to move along γ through infinitesimal steps and after each step to perform orthogonal projection of the vector thus obtained onto $T_{\gamma(t)}M^2$. It appears that such

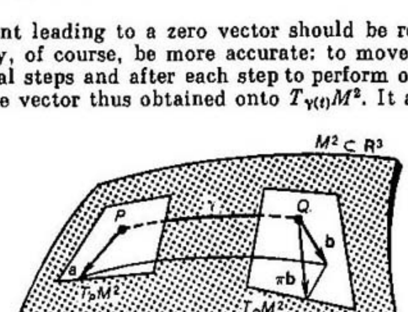

Figure 5.12

an operation can be defined correctly, and it does define a "parallel displacement" on $M^2 \subset \mathbf{R}^3$. We omit the details, for this procedure employs the embedding of M^2 in $\mathbf{R}^3$. It would be desirable to elaborate a general concept of parallel displacement which would not rely upon a particular embedding of M^n in $\mathbf{R}^N$.

It is worth noting that in order to define parallel displacement, we have to fix a smooth curve along which the operation is per-

formed. Such a curve arises quite naturally; otherwise, the phrase "to translate a vector from a point P to a point Q" loses its meaning. Roughly speaking, to "translate" a vector, one has to displace it

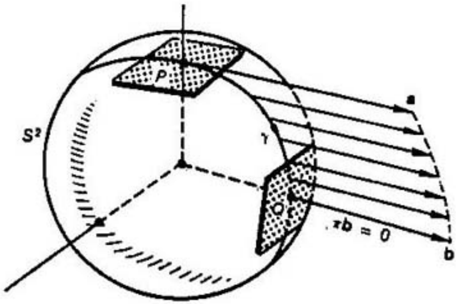

Figure 5.13

along M^n, thereby tracing a trajectory. Imagine that we are displacing a vector **a** in some M^n. It is not clear a priori whether the result of this displasiment depends on the path along which the vector is displaced. It may depend, the parallel displacement along the central line of a Möbius band being an example (see Fig. 5.14). In this case the result of the operation depends on the path because M^2 is non-orientable. With an orientable manifold, too, it is not obvious that the parallel displacement is path-independent.

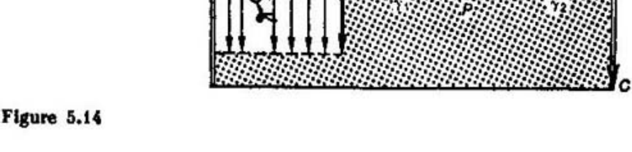

Figure 5.14

5.4.2. THE EQUATION OF PARALLEL DISPLACEMENT

If we recall the definition of the directional derivative, we see that this definition relies upon the possibility of comparing the values of a tensor field at nearby points in the same coordinate system, so that once covariant differentiation has been defined, we can define infinitesimal shifts. Thus, we shall define parallel displacement on M^n, proceeding from the operation ∇.

Let P and Q be arbitrary points on M^n connected by a smooth (or piecewise-smooth) trajectory $\gamma(t)$ such that $\gamma(0) = P$ and $\gamma(1) = Q$, and let $\dot{\gamma}$ be the velocity field along $\gamma(t)$. The components of this field in the coordinate system $x^1, \ldots, x^n$ will be denoted by $\{\xi^k\}$, $1 \leqslant k \leqslant n$. Let, on M^n, there be given an affine connection defined in the coordinate system by the set of partial derivatives $\nabla = \{\nabla_k\}$. We define the covariant derivative of a tensor field $T = \{T^{(\alpha)}_{(\beta)}\}$ along the vector field $\dot{\gamma}$ by the formula $\nabla_{\dot{\gamma}}(T) = \{\nabla_{\dot{\gamma}} T^{(\alpha)}_{(\beta)}\}$, $\nabla_{\dot{\gamma}} T^{(\alpha)}_{(\beta)} = \xi^k \nabla_k (T^{(\alpha)}_{(\beta)})$ or, conventionally, $\nabla_{\dot{\gamma}} = \xi^k \nabla_k$. This operation is called *covariant differentiation along a curve*.

Definition. Let $\gamma(t) \subset M^n$ be a smooth curve and let a field $\mathbf{T} = \{T^i\}$ be given along this curve. This field is said to be *parallel along $\gamma(t)$ relative to the connection ∇ if $\nabla_{\dot{\gamma}} \gamma(\mathbf{T}) \equiv 0$.*

By virtue of the definition of $\nabla_{\dot{\gamma}} \xi^k \nabla_k$, we may say that the field $\mathbf{T}$ which is "parallel along $\gamma(t)$" has covariantly constant coordinates along $\gamma(t)$. This is an analogue of the Euclidean case, since, for $\nabla = \left\{\nabla_k = \frac{\partial}{\partial x^k}\right\}$ and $M^n = \mathbf{R}^n$ our definition of a parallel field is reduced to the ordinary definition of parallel displacement.

Let us fix coordinates $x^1, \ldots, x^n$ and write out the condition that the field $\mathbf{T}$ is parallel. We have

$$\nabla_{\dot{\gamma}}(T^i) = \xi^k \nabla_k T^i = 0, \quad \xi^k = \frac{dx^k(t)}{dt},$$

$$\gamma(t) = (x^1(t), \ldots, x^n(t)), \quad \frac{dx^k}{dt}\left(\frac{\partial T^i}{\partial x^k} + T^p \Gamma^i_{pk}\right) = 0,$$

$$\frac{dx^k}{dt}\frac{\partial T^i}{\partial x^k} + T^p \frac{dx^k}{dt}\Gamma^i_{pk} = \frac{dT^i}{dt} + T^p \Gamma^i_{pk}\frac{dx^k}{dt} = 0.$$

Definition. The equation $\frac{dT^i}{dt} + T^p \Gamma^i_{pk} \frac{dx^k}{dt} = 0$ is called the *equation of parallel displacement along a curve $\gamma(t)$.*

For different curves γ we obtain different equations of parallel displacement. This equation (a system of n first-order equations) permits one to find the parallel field components T^i. Since $\gamma(t)$

is given, the functions $\frac{dx^k(t)}{dt}$ are known. We now turn to a particular problem of parallel displacement of a vector.

Let $\gamma(t)$ be a smooth curve from P to Q and let $\mathbf{a} = \{a^i\} \in T_P M^n$ be a vector defined at P. Our task is to construct at the point Q a new vector $\mathbf{b} \in T_Q M^n$ which we might call a "vector parallel to $\mathbf{a}$". Let us consider the equation of parallel displacement along γ. In this equation the functions Γ^i_{pk} and $\frac{dx^i(t)}{dt}$ are assumed to be known and it is required to find the unknown functions $\{T^i(t)\}$ (the components of the parallel field $\mathbf{T}(t)$) such that the condition $T^i(0) = a^i$ be satisfied at the initial moment. As is known from the theory of ordinary differential equations, the system $\frac{dT^i}{dt} + T^p \Gamma^i_{pk} \frac{dx^k}{dt} = 0$ has a solution which is unique and can be continued up to Q.

Definition. The vector $\mathbf{b} = \mathbf{T}(1) \in T_Q M^n$, which arises at Q, is called *parallel* to the vector $\mathbf{a} \in T_P M^n$ along the curve $\gamma(t)$; $\gamma(0) = P$, $\gamma(1) = Q$.

Apparently, $\mathbf{b}$ depends, in general, on the curve γ along which it is displaced. If $M^n = \mathbf{R}^n$, the vector $\mathbf{b}$ is parallel to $\mathbf{a}$ in the ordinary sense, provided for ∇ we take the Euclidean connection, i.e. put $\Gamma^i_{jk} \equiv 0$.

We now consider the properties of parallel displacement on a Riemannian manifold M^n. Let ∇ be a Riemannian connection.

Theorem 1. *Let* $\mathbf{a}, \mathbf{b} \in T_P M^n$ *be arbitrary vectors and let* $\gamma(t)$ *be a smooth curve from* P *to* Q. *Consider the parallel displacement of* $\mathbf{a}$ *and* $\mathbf{b}$ *along* $\gamma(t)$. *This operation preserves the scalar product of vectors, i.e. if* $\mathbf{a}(t)$ *and* $\mathbf{b}(t)$ *are parallel fields along* $\gamma(t)$, $\mathbf{a}(0) = \mathbf{a}$, $\mathbf{b}(0) = \mathbf{b}$, *then* $\frac{d}{dt}\langle \mathbf{a}(t), \mathbf{b}(t)\rangle \equiv 0$, *where* $\langle\ ,\ \rangle$ *is the scalar product in* $T_{\gamma(t)} M^n$ *generated by* g_{ij}.

Proof. We include $\mathbf{a}$ and $\mathbf{b}$ in the parallel fields $\mathbf{T} = \mathbf{a}(t)$ and $\mathbf{R} = \mathbf{b}(t)$, $\mathbf{T}(0) = \mathbf{a}$, $\mathbf{R}(0) = \mathbf{b}$. Let us consider the function $f(t) = \langle \mathbf{T}, \mathbf{R}\rangle_{\gamma(t)} = (g_{ij} T^i R^j)(t)$, i.e. the scalar product along $\gamma(t)$. Differentiation yields

$$\begin{aligned} \frac{df(t)}{dt} &= \nabla_{\dot\gamma} f(t) = \xi^k \nabla_k f \\ &= \xi^k \nabla_k (g_{ij} T^i R^j) = \xi^k g_{ij} \nabla_k (T^i R^j) \\ &= \xi^k g_{ij} (\nabla_k T^i) R^j + \xi^k g_{ij} T^i (\nabla_k R^j) \\ &= g_{ij} R^j (\xi^k \nabla_k T^i) + g_{ij} T^i (\xi^k \nabla_k R^j) \\ &= g_{ij} R^j (\nabla_{\dot\gamma} T^i) + g_{ij} T^i (\nabla_{\dot\gamma} R^j) \equiv 0, \end{aligned}$$

because $\nabla_{\dot\gamma}(\mathbf{T}) = \nabla_{\dot\gamma}(\mathbf{R}) = 0$. The theorem is proved.

The converse is also true: *let, on a Riemannian manifold M^n, there be given a symmetric affine connection in which the parallel displacement along any curve preserves the scalar product, then this connection is Riemannian.* Indeed, from the proof of Theorem 1 we obtain: $\xi^k T^i R^j \ (\nabla_k g_{ij}) \equiv 0$, i.e. $\nabla_k g_{ij} \equiv 0$.

Thus far, we have considered parallel displacement along a smooth curve, it is a simple matter to define such a displacement along a piece-

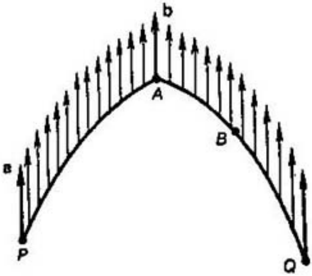

Figure 5.15

wise-smooth curve. Indeed, let $\gamma(t)$ have a cusp such that the curve is smooth and has a non-zero velocity vector on the right and left of the cusp (Fig. 5.15). Displacing vector **a** along $\gamma(t)$, we approach the point A to obtain at this point vector **b** parallel to **a**. Take **b** as the initial position of a new parallel field along AB and repeat the procedure.

5.4.3. GEODESICS

Definition. Let M^n be provided with an affine connection (the metric may not exist). A smooth curve $\gamma(t)$ is called a *geodesic in the connection* ∇ if $\nabla_{\dot{\gamma}}(\dot{\gamma}) = 0$, where $\dot{\gamma}$ is the velocity field of the trajectory $\gamma(t)$.

In other words, a geodesic is a trajectory along which parallel displacement of its velocity vector generates a velocity field along the entire curve: namely, a velocity vector turns into a velocity vector, always remaining tangent to the trajectory. Let us derive the equation of a geodesic.

We have

$$0 = \nabla_{\dot{\gamma}}(\dot{\gamma}) = \left\{\frac{dx^k}{dt}\left(\nabla_k \frac{dx^i}{dt}\right) = 0\right\},$$

if $T^i = \frac{dx^i}{dt}$, then $\frac{dT}{dt} + T^\alpha T^k \Gamma^i_{\alpha k} = 0$ or

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{\alpha k}\frac{dx^\alpha}{dt}\frac{dx^k}{dt} = 0. \tag{G}$$

Equations (G) are called the *equations of geodesics*. Their solutions are sets of functions $x^1(t), \ldots, x^n(t)$ which define the trajectory $\gamma(t)$, a geodesic. System (G) is a system of n second-order ordinary differential equations and its solution is uniquely determined by the initial values $x^i(0) = P^i$, where $P = (P^1, \ldots, P^n)$, $\frac{dx^i(0)}{dt} = a^i$, $\mathbf{a} \in T_P M^n$ ($2n$ constants in all, n constants define the position of the point P through which the solution passes, and the other n constants define the velocity vector at this point). The familiar theorems of the theory of ordinary differential equations imply the following statement.

Statement 1. *Let $P \in M^n$ and $\mathbf{a} \in T_P M^n$. Then there exists a geodesic $\gamma(t)$ such that $\gamma(0) = P$ and $\dot{\gamma}(0) = \mathbf{a}$, and this geodesic is unique.*

Proof. We introduce coordinates $x^1, \ldots, x^n$ in the neighbourhood of P. To find the geodesic, we should find a solution of system (G); the existence and uniqueness of the solution is ensured by the theorems of the theory of ordinary differential equations.

Corollary. Two geodesics, which touch each other at a point, coincide.

Let us consider a Riemannian manifold M^n and a Riemannian connection ∇. We consider also geodesics generated by ∇ and examine a parallel displacement along these geodesics. Let γ be a geodesic, $\dot{\gamma}$ the velocity field, and $\mathbf{T}$ a vector field parallel along γ. Then at each point of $\gamma(t)$ there is valid the number $\cos\alpha(t) = \langle \mathbf{T}, \dot{\gamma}\rangle / |\mathbf{T}| \times |\dot{\gamma}|$, where $\alpha(t)$ is the angle between the vectors $\mathbf{T}$ and $\dot{\gamma}$.

Lemma 1. *The parallel displacement of the vector $\mathbf{T}$ along the geodesic γ preserves the angle α: $\alpha(t) \equiv$ const.*

Proof. According to Theorem 1, all pairwise scalar products are preserved, i.e. in the identity $\langle \mathbf{T}, \dot{\gamma}\rangle = |\mathbf{T}|\,|\dot{\gamma}|\cos\alpha(t)$ both the left-hand side and the magnitudes $|\mathbf{T}|$, $|\dot{\gamma}|$ are preserved.

In a multi-dimensional case this condition is insufficient to uniquely define parallel displacement along a geodesic. For M^2, however, we obtain, by virtue of Lemma 1, the following: let γ be a geo-

desic and let $\mathbf{a} \in T_P M^2$ (Fig. 5.16), then the parallel field $\mathbf{T}(t)$, for which $\mathbf{T}(0) = \mathbf{a}$, is formed by the vectors $\mathbf{T}(t)$ having the same length as $\mathbf{T}(0) = \mathbf{a}$ and subtending with the vector $\dot{\gamma}(t)$ the same angle α as the angle between $\mathbf{a}$ and $\dot{\gamma}(0)$. Since parallel displacement along a geodesic has already been defined, we can define parallel displacement along any piecewise-smooth curve γ. To this end,

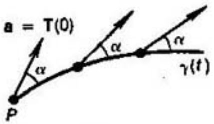

Figure 5.16

γ should be approximated by a broken geodesic (composed of smooth geodesic segments), and the parallel displacement should be effected along each of these smooth segments, keeping the angle α constant. Under parallel displacement along a trajectory which is not a geodesic, the angle between the displaced vector and the trajectory velocity vector may, in general, vary. Let us consider several examples.

(1) If $M^2 = \mathbf{R}^2$, then parallel displacement along a smooth curve is carried out in accordance with the general rule: namely, the parallel field has constant components relative to a Cartesian coordinate system.

(2) Let M^2 be a right circular cone in $\mathbf{R}^3$ with the vertex angle θ (see Fig. 5.17). To deal with a smooth manifold in $\mathbf{R}^3$, we assume

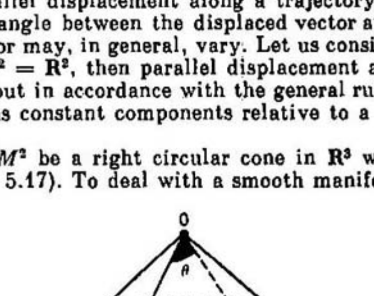

Figure 5.17

that the vertex is "punctured". Let $\gamma(t)$ be the section of the cone by a plane orthogonal to the axis, $OP = r$, and the vector $\mathbf{a} \in T_P M^2$ points to the cone vertex. Displace this vector along $\gamma(t)$ until it returns to the point P and find the angle through which $\mathbf{a}$ rotates. We shall use the connection compatible with the induced metric

on the cone. Since the metric is Euclidean, the cone can be developed by cutting it along the generator. It is sufficient to determine the rotation of a under its parallel displacement on $\mathbf{R}^2$ along γ_1 (see Fig. 5.18). We have: $PS = r \sin \frac{\theta}{2}$, the length of $\gamma(t) =$

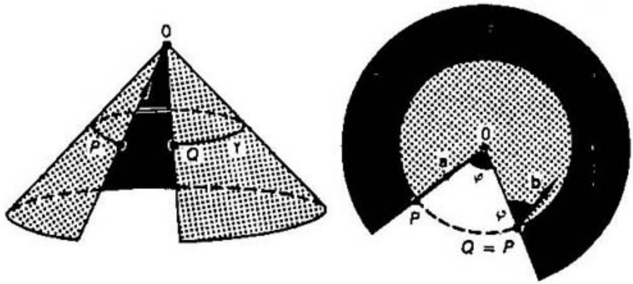

Figure 5.18

$2\pi \cdot PS = 2\pi r \sin \frac{\theta}{2}$, the length of PQ (dashed arc) $= 2\pi r - 2\pi r \times \sin \frac{\theta}{2} = 2\pi r \left(1 - \sin \frac{\theta}{2}\right) = r\varphi$, i.e. $\varphi = 2\pi \left(1 - \sin \frac{\theta}{2}\right)$. Thus, the rotation angle φ is equal to $2\pi \left(1 - \sin \frac{\theta}{2}\right)$. Here we have used $\Gamma^i_{jk} \equiv 0$. For M^2 with non-Euclidean metric the procedure is more complicated and requires the calculation of the Christoffel symbols. Let us discuss the geometric meaning of these coefficients from the standpoint of parallel displacement. Writing $\nabla_{\partial_\alpha}(\partial\beta)$, where ∂_α and ∂_β are coordinate vector fields, in the coordinates $x^1, \ldots, x^n$ on M^n, we see that the tensor $\nabla_{\partial_\alpha}(\partial_\beta)$ is again a vector field.

Lemma 2. *The identity* $\nabla_{\partial_\alpha}(\partial_\beta) = \Gamma^k_{\beta\alpha}\partial_k$ *s valid.*

Proof. We have

$$\Delta_X(Y) = Z, \quad \nabla_X(Y^i) = X^\alpha \nabla_\alpha (Y^i) = Z^i, \quad [\nabla_{\partial_\alpha}(\partial_\beta)]^k = a^q_\alpha \nabla_q (T^k_\beta),$$

where

$$\partial_\alpha = \{a^q_\alpha = \delta^q_\alpha\}, \quad \partial_\beta = \{T^k_\beta = \delta^k_\beta\},$$

$$a^q_\alpha \nabla_q (T^k_\beta) = a^q_\alpha \left(\frac{\partial}{\partial x^q} T^k_\beta + T^\omega_\beta \Gamma^k_{\omega q}\right) = a^q_\alpha T^\omega_\beta \Gamma^k_{\omega q} = \delta^q_\alpha \delta^\omega_\beta \Gamma^k_{\omega q} = \Gamma^k_{\beta\alpha},$$

i.e.

$$\nabla_{\partial_\alpha}(\partial_\beta) = \Gamma^k_{\beta\alpha}\partial_k.$$

The lemma is proved.

This statement can be interpreted as follows. Let us consider a frame $\partial_1, \ldots, \partial_n$ at a point P and make an infinitesimal translation of the vector ∂_β along ∂_α; in this case ∂_β experiences "rotation", and $\Gamma^k_{\beta\alpha}$ are the coefficients of the decomposition of ∂_β in $\partial_1, \ldots, \partial_n$.

(3) Let $M^n = \mathbf{R}^n$ be referred to Cartesian coordinates $x^1, \ldots, x^n$, then $\Gamma^i_{jk} \equiv 0$ (the connection is Riemannian). The equations of geodesics are of the form

$$\frac{d^2x^\alpha}{dt^2} = 0, \quad 1 \leqslant \alpha \leqslant n, \text{ i.e. } x^\alpha = a^\alpha t + b^\alpha,$$

where $\{a^\alpha, b^\alpha\} = \text{const}$. Thus, the geodesics are straight lines and only such lines.

(4) Let $M^2 = S^2$ in the standard metric. Choose on S^2 spherical coordinates (θ, φ) in which $ds^2 = d\theta^2 + \sin^2\theta\, d\varphi^2$, here the north pole is $\theta = 0$. To find geodesics, it is required to calculate Γ^i_{jk}. Since $(g_{ij}) = \begin{pmatrix} 1 & 1 \\ 0 & \sin^2\theta \end{pmatrix}$, we have (verify!) $\Gamma^1_{22} = -\frac{1}{2}\sin 2\theta$, $\Gamma^2_{12} = \cot\theta$, $\Gamma^i_{jk} = 0$ for all other combinations of indices (i, j, k); here $x^1 = \theta$, $x^2 = \varphi$, $g_{11} = 1$, $g_{12} = g_{21} = 0$, $g_{22} = \sin^2\theta$. It follows that the equations of geodesics are of the form $\frac{d^2\theta}{dt^2} - \frac{1}{2}\sin 2\theta\left(\frac{d\varphi}{dt}\right)^2 = 0$, $\frac{d^2\varphi}{dt^2} + \cot\theta\, \frac{d\varphi}{dt}\frac{d\theta}{dt} = 0$. One of the solutions is $\varphi = \text{const}$, $\theta = t$, i.e. a meridian emerging from the north pole. Thus, one of the sections of a sphere by a plane through the centre is a geodesic if by the parameter θ we mean the arc length t.

Proposition 1. *Let S^2 be endowed with the standard metric. Then a curve on the sphere S^2 is a geodesic of the Riemannian connection if and only if it is a plane section of the sphere through its centre.*

Proof. We first prove that any central plane section of S^2 is a geodesic (with respect to the natural parameter). This fact has already been established for the meridian γ_0 (see above). Let us consider an arbitrary "equator" γ, i.e. a plane central section. Since each equator is uniquely determined by an orthogonal straight line to the plane defining the equator, there always exists a rotation which transforms γ into γ_0.

Lemma 3. *Let $f\colon M^n \to M^n$ be the isometry of a Riemannian manifold M^n, and let γ be a geodesic of the Riemannian connection. Then the image of γ under the isometry f is also a geodesic.*

Proof. Apparently, isometry preserves the Riemannian connection and, therefore, it also preserves the equations of geodesics, i.e. f transforms a solution of the system into another solution, which is what was required.

Returning to the proof of Proposition 1, we see that γ is a geodesic. Conversely, let γ be a geodesic on S^2. Consider the velocity vector $\dot{\gamma}$ at an arbitrary point of γ and draw through this point the equator along the velocity vector (one and only one equator passes through any point on S^2 in a given direction) (see Fig. 5.19). Since γ and the

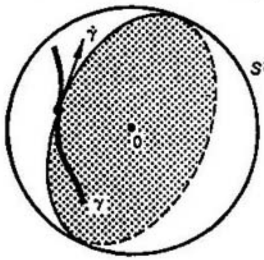

Figure 5.19

equator are solutions of the same system and since these solutions touch each other, they coincide.

(5) Let $M^2 = L_2$ be a Lobachevskian plane referred to the standard metric $ds^2 = d\chi^2 + \sinh^2\chi d\varphi^2$. We now find geodesics of the Riemannian connection. The metric takes this form in polar coordinates (χ, φ) on a two-dimensional plane. Let us consider the Poincaré model with the metric $(1 - r^2)^{-2}(dr^2 + r^2d\varphi^2)$.

Proposition 2. *All circular arcs meeting the absolute at right angles (in particular, all the diameters) are geodesics of the Lobachevskian metric in the Poincaré model. Any geodesic of the Lobachevskian metric is of the above form.*

Let us derive the equations of geodesics. Since the metric of L_2 is obtained from the metric of S^2 by substituting hyperbolic functions for trigonometric ones, we have: $\Gamma^1_{22} = -\frac{1}{2}\sinh 2\chi$; $\Gamma^2_{12} = \coth\chi$, $\Gamma^i_{jk} = 0$ for the remaining combinations of indices (i, j, k), here $x^1 = \chi$, $x^2 = \varphi$. It follows that $\frac{d^2\chi}{dt^2} - \frac{1}{2}\sinh(2\chi)\cdot\left(\frac{d\varphi}{dt}\right)^2 = 0$, $\frac{d^2\varphi}{dt^2} + \coth\chi\cdot\frac{d\varphi}{dt}\cdot\frac{d\chi}{dt} = 0$. One of the solutions is $\varphi = \text{const}$, $\chi = t$, i.e. a straight line through the point 0 on the plane. Since (χ, φ) are also polar coordinates for the right-hand sheet of the hyperboloid (a pseudosphere of imaginary radius), the stereographic projection maps the straight line $\varphi = \varphi_0$, $\chi = t$ into one of the diameters

of the unit circle. Thus, we have proved that the diameter γ_0 on the Poincaré model is a geodesic.

Let us now prove that any circular arc orthogonal to the absolute is a geodesic. We shall use Lemma 3. It is required to prove that any such circle is mapped under an isometry into γ_0. To this end, we go over to the model on the upper half-plane with the metric $\frac{dx^2+dy^2}{y^2}$. We recall that there exists a homographic transformation (an isometry) which sends a unit circle into the upper half-plane. In this case the circumference is transformed into the real axis and the diameter γ_0 into a straight line orthogonal to the real axis (Fig. 5.20).

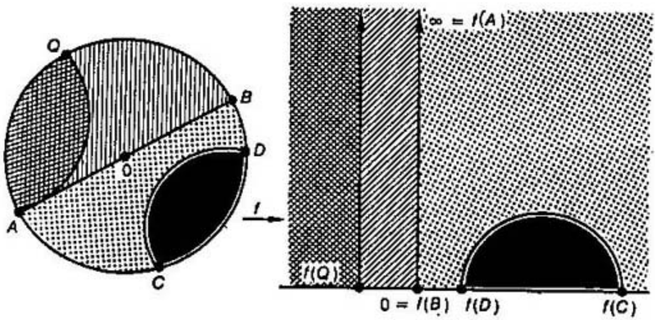

Figure 5.20

The y-axis on the upper half-plane may be regarded as a geodesic, for it is the image of the diameter γ_0 under an isometry. Thus, any straight line orthogonal to the real axis is a geodesic, since the shift $z \to z + \alpha$, $\alpha \in \mathbf{R}$, is an isometry. It follows that any circular arc meeting the x-axis at right angles is a geodesic, for it can be transformed into the y-axis under a homographic mapping: first, the shift $z \to z + \alpha$, $\alpha \in \mathbf{R}$, and then $z \to \frac{-z}{z-a}$, $a \in \mathbf{R}$ (see Fig. 5.21). We have thus proved that all straight lines orthogonal to the real axis and all circular arcs meeting the x-axis at right angles are geodesics. Let γ be an arbitrary geodesic (on the upper half-plane), it is required to prove that this geodesic coincides either with a straight line orthogonal to the x-axis or with a circle orthogonal to the absolute.

Choose a point P on γ and consider the vector $\dot{\gamma}$. Draw through P a circular arc which meets the absolute at right angles and has the same velocity vector (see Fig. 5.22). As we have already demonstrated, γ coincides with this arc.

As an example, we shall consider the parallel displacement of a vector on a Lobachevskian plane along the trajectory $\gamma(t)$ defined on the upper half-plane by the equation $y = y_0 = \text{const}$, i.e. along

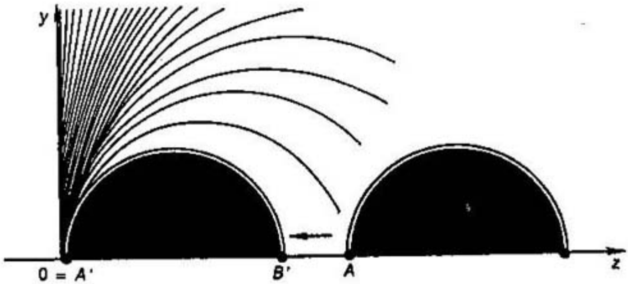

Figure 5.21

the straight line parallel to the x-axis. This trajectory is not a geodesic and, therefore, parallel displacement along this trajectory does not preserve the tangent velocity field. Approximate γ by a broken

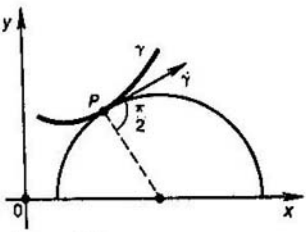

Figure 5.22

geodesic (see Fig. 5.23). A qualitative picture of parallel displacement of a is shown in Fig. 5.24. The vector a rotates about its origin.

Remark. Geodesics on a Lobachevskian plane can be found in a more elementary way: by integrating explicitly the equations of geodesics. It is convenient to perform integration on the upper half-plane. The Christoffel symbols are of the form (verify!): $\Gamma^1_{12} = \frac{1}{y}$, $\Gamma^2_{11} = -\frac{1}{y}$, $\Gamma^2_{22} = \frac{1}{y}$, the remaining symbols vanish. The equa-

tions of geodesics are $\ddot{x} = \frac{2\dot{x}\dot{y}}{y}$, $\ddot{y} = \frac{\dot{y}^2 - \dot{x}^2}{y}$, whence

$$\frac{d^2y}{dx^2} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\dot{x}^3} = \frac{\frac{\dot{y}^2 - \dot{x}^2}{y}\cdot\dot{x} - \frac{2\dot{x}\dot{y}}{y}\cdot\dot{y}}{(\dot{x})^3}$$

$$= -\frac{1}{y}\left(\frac{\dot{y}^2}{\dot{x}^2} + 1\right) = -\frac{1}{y}(y_x'^2 + 1),$$

$$y'' = -\frac{1}{y}(y_x'^2 + 1), \quad yy'' + y'^2 = -1, \quad (yy')' = -1,$$

$$yy' = -x + C, \quad y\,dy = (-x + C)\,dx, \quad \frac{y^2}{2} = -\frac{x^2}{2} + Cx + \frac{D}{2},$$

$$x^2 - 2Cx + y^2 = D, \quad (x - C)^2 + y^2 = C^2 + D.$$

In this calculation it was assumed that $\dot{x} \neq 0$. If $\dot{x} = 0$, we obtain straight lines orthogonal to the real axis, and if $x \neq 0$ we have, evidently, circles orthogonal to the absolute.

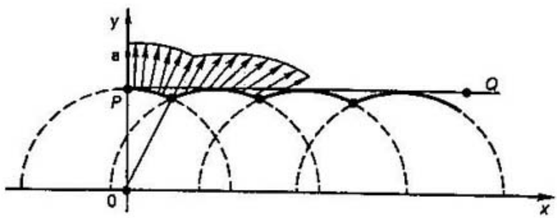

Figure 5.23

Problem. Displace a vector **a** along a plane (non-central) section of S^2 (see Fig. 5.25).

(6) Let $M^2 = T^2$ be a two-dimensional torus. It can be transformed into a Riemannian manifold with a locally Euclidean metric. On the circles $S^1(\varphi)$ and $S^1(\psi)$ we introduce the coordinates φ and ψ, respectively; then (φ, ψ), $0 \leqslant \varphi, \psi \leqslant 2\pi$, are coordinates on the torus, and the metric (in these coordinates) takes the form $d\varphi^2 + d\psi^2$. This metric may be considered as induced by the Euclidean metric from $\mathbf{R}^4$ under the embedding of T^2 in $\mathbf{R}^4 \cong \mathbf{C}^2$ given by $f(\varphi, \psi) \to (e^{i\varphi}, e^{i\psi}) \in \mathbf{C}^2$. Since $ds^2(\mathbf{R}^4) = dz\,d\bar{z} + dw\,d\bar{w}$, we have $ds^2(T^2) = d\varphi^2 + d\psi^2$.

Thus, since in the Euclidean metric $\Gamma^i_{jk} = 0$, only the images of straight lines on $\mathbf{R}^2(\varphi, \psi)$ under the factorization h: $\mathbf{R}^2(\varphi, \psi) \to T^2(\varphi, \psi)$, $h(\varphi, \psi) = (\varphi, \psi)\cdot\text{mod } 2\pi$, are geodesics on a torus, i.e.

$T^2 = \mathbf{R}^2/\mathbf{Z} \oplus \mathbf{Z}$ (see Fig. 5.26). The are no other geodesics on the torus. The geodesics are divided into two classes: closed and unclosed. It is convenient to depict geodesics on

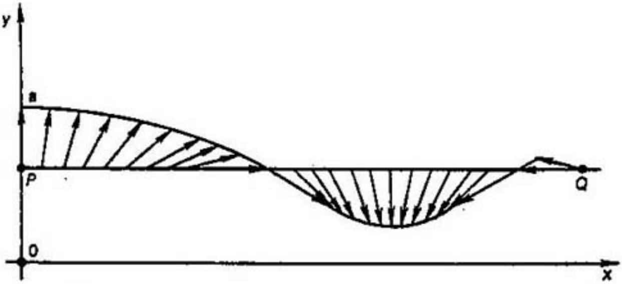

Figure 5.24

a torus as straight lines on $\mathbf{R}^2(\varphi, \psi)$ with a fixed lattice $\mathbf{Z} \oplus \mathbf{Z} = (2\pi m, 2\pi n)$, $m, n \in \mathbf{Z}$. Let us consider a sheaf of straight lines emerging from the point 0 on $\mathbf{R}^2(\varphi, \psi)$ and find geodesics which are

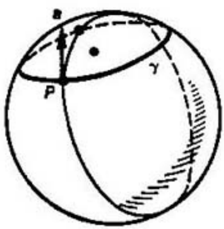

Figure 5.25

the images of these straight lines. Apparently, the geodesic through the point (0, 0) on a torus is closed if and only if the inverse image of this geodesic (a straight line) meets some "integer" point $(2\pi m, 2\pi n)$, $m, n \in \mathbf{Z}$. It follows that the geodesic on T^2 through (0, 0) is unclosed (homeomorphic to a straight line) if and only if the corresponding straight line does not contain "integer" points except for (0, 0). This can also be expressed in terms of the slope of the straight line l to the x-axis: namely, a geodesic is closed (homeomorphic to a circle) if and only if $\tan\alpha = X/Y$ is rational (here (X, Y) are the coordinates of the direction vector of the straight line l (see Fig. 5.27)); and a geodesic is unclosed if $\tan\alpha$ is irrational. In Fig. 5.27 the straight line passes through the point $(2\pi \cdot 3, 2\pi \cdot 2)$; after factori-

zation we obtain on the square $0 \leqslant \varphi \leqslant 2\pi$, $0 \leqslant \psi \leqslant 2\pi$ a collection of segments—the images of this straight line (see Fig. 5.28). Figure 5.28 also shows the trajectory which arises on the torus after

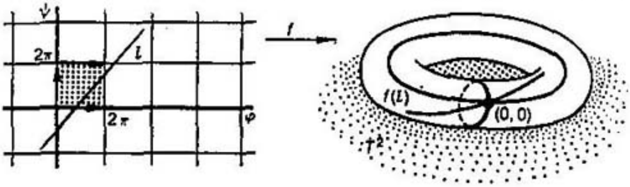

Figure 5.26

the fundamental polygon is glued in accordance with the operation of the group $\mathbf{Z} \oplus \mathbf{Z}$. Figure 5.29 illustrates the geodesic with rational ($\tan \alpha = n/m$) and irrational slope. In the latter case the straight

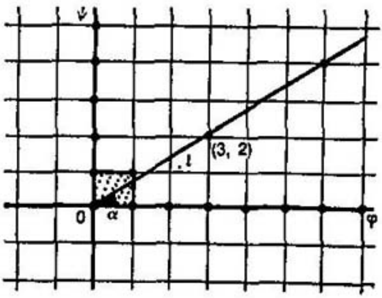

Figure 5.27

segments are everywhere dense on the square (under factorization the segment of the straight line l passes as close as possible to any point). The trajectory thus obtained on the torus induces an irrational torus winding, the entire torus being the closure of the trajectory (see Fig. 5.30).

We now apply the concept of geodesics to particular geometric problems.

Theorem 2. *Let M^2 be one of the following manifolds*: (a) $\mathbf{R}^2$, (b) S^2, (c) L_2 (*Lobachevskian plane*) *provided with standard metrics and let* $\mathfrak{G} = \operatorname{Iso} M^2$ *be the group of all isometries of* M^2. *Then each mapping* $g \in \mathfrak{G}$ *is defined by three continuous parameters, i.e.* $\dim \mathfrak{G} = 3$.

By $\mathfrak{G}$ we mean the complete isometry group, i.e. the group of diffeomorphisms preserving the metric. The group $\mathfrak{G} = \text{Iso } M^n$ can be defined for any smooth Riemannian manifold. This group can be transformed into a topological space: namely, two mappings g_1, g_2

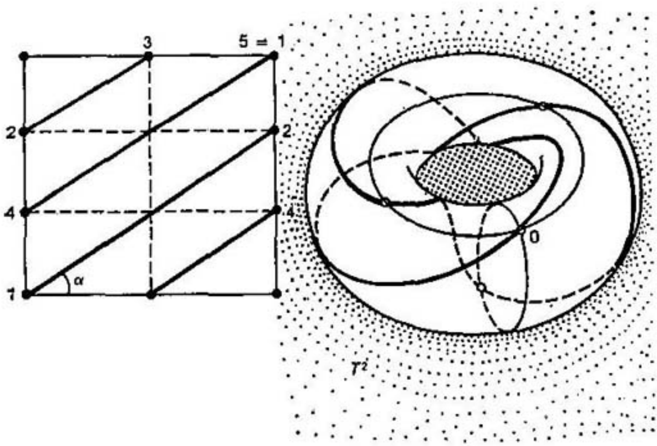

Figure 5.28

are assumed to be close if they are close as diffeomorphisms of M^n, i.e. for all points $x \in M^n$ the points $g_1(x)$ and $g_2(x)$, $g_1, g_2 \in \mathfrak{G}$, are close.

Theorem 2 is a particular case of a more general statement which is presented here without proof.

Let M^n be a compact, smooth, Riemannian, connected, closed manifold and let $\mathfrak{G} = \text{Iso } M^n$, then $\dim \mathfrak{G} \leqslant \frac{n(n+1)}{2}$, *i.e. each mapping $g \in \mathfrak{G}$ is defined by not more than $\frac{n(n+1)}{2}$ continuous parameters.*

This general theorem will not be used in the sequel. Although we shall prove Theorem 2 only for the three manifolds mentioned above, the reasoning remains true for an arbitrary manifold, in the course of the proof we specify the places where a particular form of M^2 is employed.

Proof. Let $x_0 \in M^n$ and let $H(x_0) \subset \mathfrak{G}$ be the set of isometries such that x_0 remains fixed. Evidently, $H(x_0)$ is a subgroup. It is called a *stability subgroup of the point* x_0; for distinct x_1 and x_2 the subgroups $H(x_1)$ and $H(x_2)$ are, in general, distinct. Let $h \in H(x_0)$.

Since $h(x_0) = x_0$, we have $dh(x_0)$: $T_{x_0}M^n \subset T_{x_0}M^n$. Construct the mapping λ: $H(x_0) \to \mathfrak{G}(n, \mathbf{R})$, setting $\lambda(h) = dh(x_0)$. Apparently, $dh \in \mathfrak{G}L(n, \mathbf{R})$ because h is a diffeomorphism. Furthermore, λ maps $H(x_0)$ into the subgroup $O(n) \subset \mathfrak{G}L(n, \mathbf{R})$. Indeed, we may assume

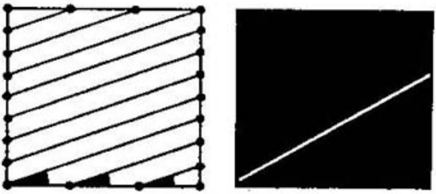

Figure 5.29. tan α is rational number on the left, and it is irrational on the right.

that the coordinates chosen in the neighbourhood of x_0 are such that $g_{ij}(x_0) = \delta_{ij}$, and then $g_{ij}(x_0)$ defines in $T_{x_0}M^n$ the Euclidean metric. Since h is an isometry, $dh(x_0)$ preserves the Euclidean scalar product in $T_{x_0}M^n$. Also, λ is a homomorphism of $H(x_0)$ in $O(n)$.

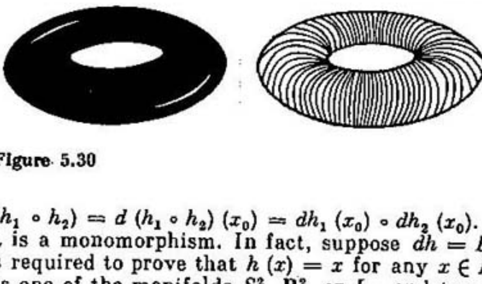

Figure 5.30

Indeed, $\lambda(h_1 \circ h_2) = d(h_1 \circ h_2)(x_0) = dh_1(x_0) \circ dh_2(x_0)$. And all the more, λ is a monomorphism. In fact, suppose $dh = E$ (identically), it is required to prove that $h(x) = x$ for any $x \in M^n$. Since $M^n = M^2$ is one of the manifolds S^2, $\mathbf{R}^2$, or L_2, and two points on each manifold can be connected by a geodesic. The statement is obvious for a plane. On S^2, according to Proposition 1, the geodesics are equators, which proves the statement. If $M^2 = L_2$, we consider the upper half-plane, the construction of the geodesic is shown in Fig. 5.31. This statement also holds true for manifolds of a more general type (the proof of this fact is omitted).

So, let $dh(x_0) = E$, connect an arbitrary point $x \in M^n$ with x_0. Let $\dot{\gamma}(0)$ be the velocity vector of $\gamma(t)$ at the point x_0. Since h is an isometry, the image of γ under h is a geodesic, and since $dh(x_0) =$

E, the geodesic $\gamma_1 = h(\gamma)$ has at x_0 the same velocity vector as γ. We have already pointed out that two tangent geodesics coincide. Since γ may be considered as the natural parameter, h does not alter this parameter along γ, so that x remains (under h) at the same distance from x_0, i.e. $h(x) = x$. And this is what was required. Thus,

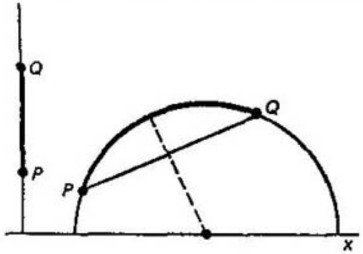

Figure 5.31

$H(x_0)$ is a closed subgroup in $O(n)$. Closeness of $H(x_0)$ follows from the fact that the isometry, which is the limit of isometries preserving x_0, also preserves x_0. For $M^2 = S^2$, $\mathbf{R}_2$, or L_2 we have, therefore, $\dim H(x_0) \leqslant \dim O(2) = 1$. For an arbitrary M^n it can be shown that $\dim H(x_0) \leqslant \dim O(n)$, i.e. $\dim H(x_0) \leqslant n(n-1)/2$. We now turn to the group $\mathfrak{G}$. **Statement**: *any isometry $g \in \mathfrak{G}$ is defined by the image of the point x_0, i.e. $g(x_0)$, and by the differential $dg(x_0)$*: $T_{x_0}M^n \to T_{g(x_0)}M^n$. Indeed, consider the correspondence $g \to (g(x_0), dg(x_0))$ and let $(g_1(x_0), dg_1(x_0)) = (g_2(x_0), dg_2(x_0))$, whence $g_1(x_0) = g_2(x_0)$ and $dg_1(x_0) = dg_2(x_0)$. Consider then $g(x) = (g_1^{-1}) \circ g_2(x)$, $x \in M^n$. We have

$$g(x_0) = (g_1^{-1}) \circ g_2(x_0) = x_0, \quad \text{i.e.} \quad g \in H(x_0),$$
$$dg(x_0) = d((g_1^{-1}) \circ g_2)(x_0) = ((dg_1)^{-1} \circ (dg_2))(x_0)$$
$$= (dg_1(x_0))^{-1} \circ (dg_2(x_0)) = E,$$

whence $g \equiv E$ on M^n, i.e. $g(x) = x$, $g_1(x) \equiv g_2(x)$. Since $g(x_0)$ is defined by n parameters and $dg(x_0)$ by not more than $\frac{n(n-1)}{2}$ parameters, g can be defined by not more than $n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$ parameters. For $M^2 = S^2$, $\mathbf{R}^2$, and L_2 we have $\dim \mathfrak{G} \leqslant 3$. On the other hand, it was shown in Chapter 4 that each of the groups Iso $\mathbf{R}^2$, Iso S^2, and Iso L_2 contains a subgroup whose transformations are also defined by three parameters. Since these sub-

groups are open and closed, we have dim Iso $M^2 = 3$ for $M^2 = S^2$, $\mathbf{R}^2$, and L_2. The theorem is proved.

Corollary. *Let* Iso $(M^n)_0$ *be the connected component of the unit element in* Iso M^n. *Then for* $M^2 = S^2$, $\mathbf{R}^2$, *and* L_2 *the group* Iso $(M^2)_0$ *coincides with the three-dimensional groups constructed in Chapter 4, i.e.* Iso $(S^2)_0 = SO(3)$, Iso $(L_2)_0 = SL(2, \mathbf{R})/\mathbf{Z}_2$, *and* Iso $(\mathbf{R}^2)_0$ *coincides with the group of all linear isometries of a plane which preserve orientation.*

Theorem 3. *Any two-dimensional smooth, compact, connected, closed manifold admits triangulation.*

Proof. Let M^2 be provided with a Riemannian metric (viz., embed M^2 in a Euclidean space) and consider geodesics on M^2. To prove the theorem, we need the following lemma.

Lemma 4. *For each point* P_0 *on a Riemannian manifold* M^n *there exist a neighbourhood* U *and a number* $\varepsilon > 0$ *such that:* (a) *any two points in* U *can be connected by only one geodesic of length smaller than* ε, (b) *this geodesic depends smoothly on the initial and terminal points.*

Proof. We first recall the theorem on ordinary differential equations; given a system $\frac{d^2\mathbf{u}}{dt^2} = \mathbf{F}\left(\mathbf{u}, \frac{d\mathbf{u}}{dt}\right)$, where $\mathbf{u} = (u^1, \ldots, u^n)$ and $\mathbf{F}$ is a set of n smooth functions defined in the neighbourhood W of the point $(\mathbf{u}_1, \mathbf{v}_1) \in \mathbf{R}^{2n}$. Then there exist a neighbourhood U of $(\mathbf{u}_1, \mathbf{v}_1)$ and a number $\varepsilon > 0$ such that for each point $(\mathbf{u}_0, \mathbf{v}_0) \in U$ the equation $\frac{d^2\mathbf{u}}{dt^2} = \mathbf{F}\left(\mathbf{u}, \frac{d\mathbf{u}}{dt}\right)$ has only one solution $t \to \mathbf{u}(t)$ defined for $|t| < \varepsilon$ and such that $\mathbf{u}(0) = u_1$, $\frac{d\mathbf{u}(0)}{dt} = \mathbf{v}_1$, this solution depends smoothly on the initial conditions. Let $P_0 \in M^n$, then, according to this statement, there exists a neighbourhood W of P_0 such that for each $P \in W$ there is valid a mapping $\exp_P$ defined as follows. Let $\mathbf{a} \in T_PM^n$ be a vector of length not larger than ε; draw along this vector a geodesic $\gamma_{\mathbf{a}}(t)$ referred to the natural parameter t and associate with $\mathbf{a}$ the point $\gamma_{\mathbf{a}}(1)$ denoted by $\exp_P(\mathbf{a})$. We obtain a smooth mapping of a ball of radius ε in M^n (differentiability follows from the existence and uniqueness theorem) (Fig. 5.32). Let us construct the mapping F: $V \to M^n \times M^n$, where V is a neighbourhood of the point $(P_0, 0)$ *in the manifold* T_*M^n, i.e. $V = \{(P, \mathbf{a}), P \in U(P_0), |\mathbf{a}| < \varepsilon\}$, and $F(P, \mathbf{a}) = (P, \exp_P(\mathbf{a}))$. The structure of a smooth manifold in T_*M^n consists of all pairs $(P, \mathbf{a})$, $P \in M^n$, $\mathbf{a} \in T_PM^n$, if $x^1, \ldots, x^n$ are coordinates in the domain $U \subset M^n$, then each $\mathbf{a} \in T_PM^n$ can uniquely be represented in the form $\mathbf{a} = t^i\partial_i$, where $\partial_i = \left.\frac{\partial}{\partial x^i}\right|_P$. The functions $(x^1, \ldots, x^n, t^1, \ldots, t^n)$ form a local coordinate system in the open set $\mathbf{R} \subset T_*M^n$. We now prove that the Jacobian of the

mapping F is non-singular at $(P_0, 0)$. Let $\{x_1^i, x_2^i\}$, $1 \leqslant i \leqslant n$, stand for coordinates in $U \times U \subset M^n \times M^n$, then

$$F_*\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial}{\partial x_1^i} + \frac{\partial}{\partial x_2^i}, \quad F_*\left(\frac{\partial}{\partial t^j}\right) = \frac{\partial}{\partial x_2^j},$$

i.e. the Jacobi matrix of F at the point $(P_0, 0)$ is $\begin{pmatrix} E & E \\ 0 & E \end{pmatrix}$, i.e. it is non-singular. The implicit function theorem implies that F maps diffeomorphically the neighbourhood W of $(P_0, 0) \in T_*M^n$ onto the neighbourhood $\mathbf{R}$ of $(P_0, P_0) \in M^n \times M^n$. The lemma is proved.

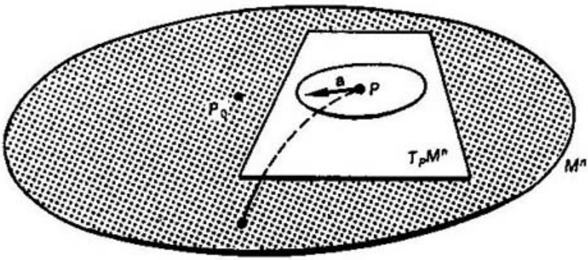

Figure 5.32

This lemma can be proved differently. It is known that a dynamical system called a *geodesic flow*, exists on T_*M^n. Let us consider a point $(P, \mathbf{a}) \in T_*M^n$, according to the existence and uniqueness theorem, only one geodesic $\gamma(t)$ emerges from P along the vector $\mathbf{a}$, i.e. there arises the velocity field $\dot{\gamma}$, and we obtain the trajectory $(P(t), \dot{\gamma}(t)) = \Gamma(t)$ in T_*M^n. These velocity vectors form the geodesic flow. The integral trajectories of this field can be extended infinitely, in particular by the distance ε (which is the same for all points in T_*M^n). This proves Lemma 4.

We now proceed with the proof of the theorem. Since M^2 is compact and closed, it can be covered with finitely many small disks. According to Lemma 4, we may assume that each of the disks is such that any two of its points can be connected by a single geodesic of length less than ε, where ε is sufficiently small. Covering M^2 with a rather dense network of points $\{P_i\}$, we can connect points inside some disk by a geodesic, thereby subdividing each disk into triangles satisfying the triangulation requirements. It is important however that the smooth subdividing curves arising in a disk should

also remain smooth in the coordinates of any other disk in which they can be located if the points they connect lie at the intersection of the disks. The subdividing curves are smooth because the solutions of the equations of geodesics are smooth, so that the triangulation can be extended until the entire M^2 is covered. The theorem is proved.

In the proof we did rely upon the fact that M^2 is two-dimensional.

5.5. CURVATURE TENSOR

5.5.1. PRELIMINARIES

Let us consider a manifold M^n (not necessarily Riemannian) with symmetric affine connection ∇. We have already proved the formula $\nabla_{\partial_\alpha}(\partial_\beta) = \Gamma^k_{\alpha\beta}\partial_k$, where ∂_α are the basis vector fields.

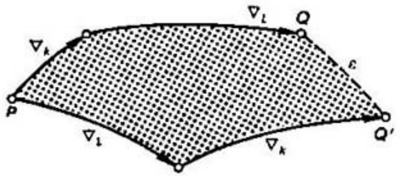

Figure 5.33

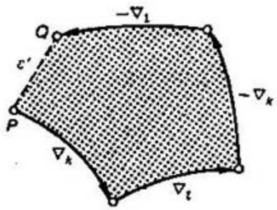

Figure 5.34

The connection ∇ defines parallel displacement, and this formula can be interpreted as follows: it describes parallel displacement (by an infinitesimal distance) along the coordinate line x^α. Let us fix a point P and consider the parallel displacements $\nabla_k\nabla_l$ and $\nabla_l\nabla_k$, where $\nabla_\alpha = \nabla_{\partial_\alpha}$ (Fig. 5.33). This is the case of displacements along the coordinate lines x^k and x^l by small distances α and β. The terminal points Q and Q' will, in general, be distinct. This effect can also be noticed with another displacement. Let us consider the motion shown in Fig. 5.34. Generally, this small "parallelogram" is non-closed, i.e. we cannot come back to the point P because M^n is "curved". This "curvature" can be measured, conventionally of course, as the difference $\nabla_k\nabla_l - \nabla_l\nabla_k = \varepsilon$. If $M^n = \mathbf{R}^n$ and we use Cartesian coordinates, then $\nabla_k\nabla_l = \nabla_l\nabla_k$; if M^n is arbitrary, this commutator need not necessarily be zero. A clear example is S^2 referred to the coordinates (θ, φ) (see Fig. 5.35). Here $\pm\alpha$ are displacements along meridians, and $\pm\beta$ are displacements along parallels.

Simple examples show that curvature may be of different types. We shall illustrate this for S^2 and for a Lobachevskian plane L_2.

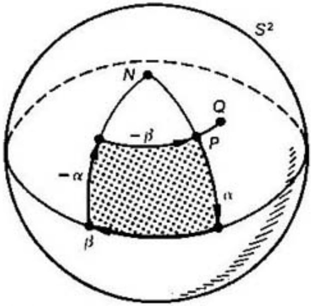

Figure 5.35

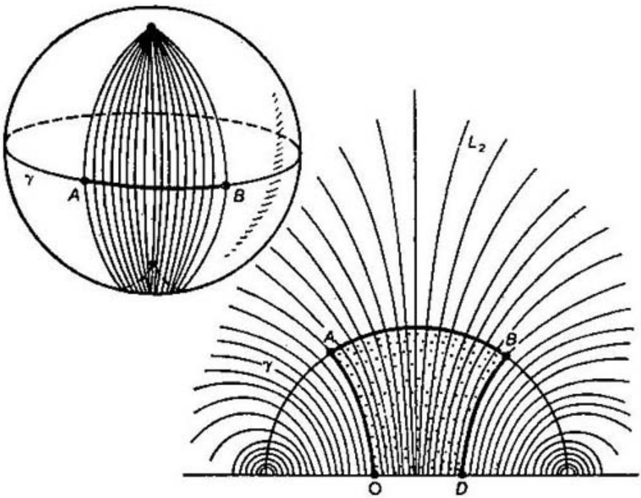

Figure 5.36

Let us consider, on S^2 and L_2, a geodesic γ and draw orthogonal geodesics from each point of its segment. We now examine the behaviour of this sheaf of trajectories orthogonal to γ. Figure 5.36

shows the qualitative picture. On S^2 the sheaf converges (along both directions) at two points: the north and south poles. On L_2 the sheaf "diverges" and the distance between the extreme geodesics

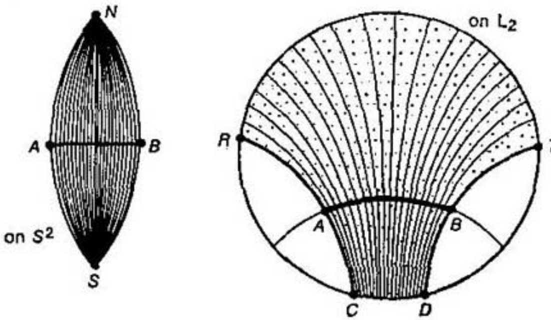

Figure 5.37

tends to infinity. It can be seen from Fig. 5.37 that on L_2 geodesics diverge on both sides of the segment AB because the arcs CD and RT have infinite lengths. Different behaviour of geodesics on S^2

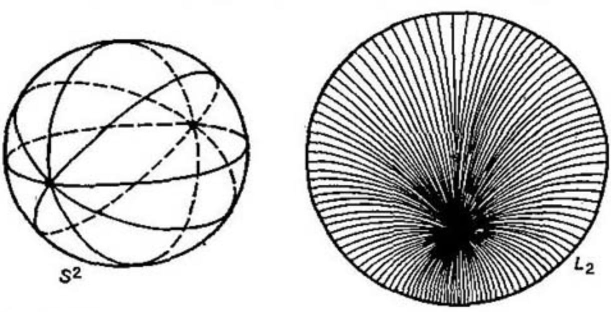

Figure 5.38

and L_2 can also be seen for a sheaf of geodesics emerging from a single point (see Fig. 5.38). We recall that S^2 and L_2 have distinct Gaussian curvatures: for S^2 the curvature is constant positive and for L_2 it is constant negative. We shall demonstrate below that the Gaussian curvature (for M^2) is closely related to the properties of the operator $\nabla_k\nabla_l - \nabla_l\nabla_k$ which measures the "curvature" of M^2.

5.5.2. COORDINATE DEFINITION OF THE CURVATURE TENSOR

Let M^n be referred to local coordinates $x^1, \ldots, x^n$ in some neighbourhood of a point P; consider $\nabla_k\nabla_l - \nabla_l\nabla_k$ and apply this operator to the field $\mathbf{T} = \{T^i\}$. The connection ∇ is symmetric, and straightforward calculation yields

$$\nabla_l T^i = \frac{\partial T^i}{\partial x^l} + T^p\Gamma^i_{pl},$$

$$\nabla_k\nabla_l(T^i) = \frac{\partial^2 T^i}{\partial x^k\,\partial x^l} + \frac{\partial T^p}{\partial x^k}\Gamma^i_{pl} + T^p\frac{\partial}{\partial x^k}(\Gamma^i_{pl}) + \nabla_l(T^p)\,\Gamma^i_{pk}$$
$$-\nabla_p(T^i)\,\Gamma^p_{kl} = \frac{\partial^2 T^i}{\partial x^k\,\partial x^l} + \frac{\partial T^p}{\partial x^k}\Gamma^i_{pl} + T^p\frac{\partial}{\partial x^k}(\Gamma^i_{pl})$$
$$+\frac{\partial T^p}{\partial x^l}\Gamma^i_{pk} + T^q\Gamma^p_{ql}\Gamma^i_{pk} - \frac{\partial T^i}{\partial x^p}\Gamma^p_{kl} - T^q\Gamma^i_{qp}\Gamma^p_{kl},$$

$$(\nabla_k\nabla_l - \nabla_l\nabla_k)\,T^i$$
$$= T^p\left[\frac{\partial}{\partial x^k}\Gamma^i_{pl} - \frac{\partial}{\partial x^l}\Gamma^i_{pk}\right] - (\Gamma^p_{kl} - \Gamma^p_{lk})\frac{\partial T^i}{\partial x^p}$$
$$+ T^q[\Gamma^p_{ql}\Gamma^i_{pk} - \Gamma^p_{pk}\Gamma^i_{pl} - \Gamma^i_{qp}\Gamma^p_{kl} + \Gamma^i_{qp}\Gamma^p_{kl}].$$

Since $\Gamma^i_{jk} = \Gamma^i_{kj}$, we have

$$(\nabla_k\nabla_l - \nabla_l\nabla_k)\,T^i = T^q\left[\frac{\partial}{\partial x^k}\Gamma^i_{ql} - \frac{\partial}{\partial x^l}\Gamma^i_{qk} + \Gamma^p_{ql}\Gamma^i_{pk} - \Gamma^p_{qk}\Gamma^i_{pl}\right] = T^q R^i_{q,kl},$$

where

$$R^i_{q,kl} = \frac{\partial\Gamma^i_{ql}}{\partial x^k} - \frac{\partial\Gamma^i_{qk}}{\partial x^l} + \Gamma^p_{ql}\Gamma^i_{pk} - \Gamma^p_{qk}\Gamma^i_{pl}.$$

Lemma 1. *The collection of $R^i_{q,kl}$ forms a tensor of rank 4.*

The *proof* is obvious, because $\nabla = \{\nabla_k\}$ is a tensor operation.

Definition. The tensor $R^i_{q,kl}$ is called the *Riemann curvature tensor of a given connection* ∇.

If $M^n = \mathbf{R}^n$, this tensor is equal to zero. Indeed, it vanishes in a Cartesian coordinate system and, therefore, it vanishes in any other coordinate system due to the tensor transformation law. There exist however M^n where $R^i_{j,kl}$ is non-zero (see below). We recall that the coordinates in which $\Gamma^i_{jk} \equiv 0$ are Euclidean for a given connection. This implies that the following lemma holds true.

Lemma 2. *Let M be provided with a symmetric affine connection. If the Riemann curvature tensor of this connection does not vanish (in a certain coordinate system), no Euclidean coordinates (in the neighbourhood of a point) can be defined on M^n.*

If such coordinates existed, Γ^i_{jk} would vanish in this coordinate system, and therefore the curvature tensor would also vanish. Thus, the curvature tensor is an "obstruction" to defining Euclidean coordinates (for a given connection).

5.5.3. INVARIANT DEFINITION OF THE CURVATURE TENSOR

In the preceding subsection we have constructed the Riemann tensor in a particular coordinate system. Now we give an invariant definition. Let **X**, **Y**, and **Z** be arbitrary smooth vector fields on M^n (with symmetric affine connection). Let us construct a "curvature operator" R which associates a new vector field with the triplet **X**, **Y**, **Z**. It is convenient to handle the fields as linear differential operators, so that below we shall write just X instead of **X**.

Definition. Put $R(X, Y) = \nabla_X\nabla_Y(Z) - \nabla_Y\nabla_X(Z) - \nabla_{[X,Y]}(Z)$. Thus, R maps $T_x \times T_x \times T_x$ into T_x, where $x \in M^n$.

Theorem 1. *The mapping R is trilinear and, therefore, defines a fourth-rank tensor.*

Proof. For linear combinations of the arguments with constant coefficients trilinearity is obvious. What is required to be proved is that a smooth function $f(x)$ can be taken out of the symbol of the operator R. If we prove this fact, R is completely defined by its values on the basis fields: $\partial_\alpha = \left(\overrightarrow{\frac{\partial}{\partial x^\alpha}}\right)$, $1 \leqslant \alpha \leqslant n$. Let us consider the mapping $(X, Y, Z) \to (X, Y, f(x) Z)$, where $f(x)$ is a smooth function. It is required to prove that $R(XY)\cdot(fZ) = f\cdot R(X, Y)\cdot Z$. We have

$$
\begin{aligned}
&\nabla_X\nabla_Y(fZ) - \nabla_Y\nabla_X(fZ) - \nabla_{[X,Y]}(fZ)\\
&= \nabla_X((\nabla_Y f) Z) + \nabla_X(f\,\nabla_Y Z)\\
&- \nabla_Y((\nabla_X f) Z) - \nabla_Y(f\,\nabla_X Z) - (\nabla_{[X,Y]} f) Z - f\nabla_{[X,Y]} Z\\
&= (\nabla_X\nabla_Y f) Z + (\nabla_Y f)\nabla_X Z + (\nabla_X f)\nabla_Y Z + f(\nabla_X\nabla_Y Z)\\
&- (\nabla_Y\nabla_X f) Z - (\nabla_X f)\nabla_Y Z - (\nabla_Y f)\nabla_X Z - f(\nabla_Y\nabla_X Z)\\
&- (\nabla_{XY-YX} f) Z - f(\nabla_{[X,Y]} Z) + \{(X(Yf) - Y(Xf) - (XY - YX) f\}\\
&+ f\{\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]} Z\} = 0 + f\cdot R(X, Y) Z,
\end{aligned}
$$

because $\nabla_X f = X(f)$. Let us verify that $R(fX, Y) Z = f\cdot R(X, Y) Z$. The following relation holds true:

$$R(fX, Y) Z = \nabla_{fX}\nabla_Y Z - \nabla_Y\nabla_{fX} Z - \nabla_{[fX,Y]} Z.$$

Evidently, $\nabla_{fX} = f\nabla_X$, since

$$(\nabla_{fX}) T = (fX)^k \nabla_k T = f\{X^k \nabla_k T\} = f(\nabla_X T).$$

Furthermore,

$$[fX, Y] = f(XY) - Y(fX) = f(XY) - (Yf)X - f(YX) = f[X, Y] - (Yf)X,$$

whence

$$R(fX, Y)Z = f(\nabla_X\nabla_Y Z) - \nabla_Y(f\nabla_X Z) - \nabla_{f[X,Y]}Z + \nabla_{(Yf)X}Z = f(\nabla_X\nabla_Y Z) - (\nabla_Y f)\nabla_X Z - f(\nabla_Y\nabla_X Z) - f\nabla_{[X,Y]}Z + (Yf)\nabla_X Z = f\cdot R(X, Y)Z + 0 = f\cdot R(X, Y)Z,$$

which is what was required. The formula $R(X, fY)Z = f\cdot R(X, Y)Z$ can be verified in a similar way. The theorem is proved.

Let us relate the invariant definition of the curvature tensor and the coordinate definition. Consider the basis fields ∂_i as differential operators and decompose X, Y, Z in these fields: $X = X^i\partial_i$, $Y = Y^j\partial_j$, $Z = Z^k\partial_k$. We obtain $R(X, Y)Z = X^iY^jZ^k\cdot\{R(\partial_i, \partial_j)\partial_k\}$; i.e. $R(X, Y)Z$ is completely determined by $R(\partial_i\partial_j)\partial_k$. Moreover,

$$R(\partial_i, \partial_j)Z = \nabla\partial_i\nabla\partial_j Z - \nabla\partial_j\nabla\partial_i Z - \nabla_{[\partial_i,\partial_j]}Z.$$

Apparently, $\nabla_{\partial_i} = \nabla_i$ (by the definition of ∇_{∂_i}), i.e. $R(\partial_i, \partial_j)Z = (\nabla_i\nabla_j - \nabla_j\nabla_i)Z - \nabla_{[\partial_i,\partial_j]}Z$. Since $[\partial_i, \partial_j] = 0$, we have $R(\partial_i, \partial_j)Z = (\nabla_i\nabla_j - \nabla_j\nabla_i)Z$. Thus, we have obtained (in a fixed coordinate system $x^1, \ldots, x^n$) the "coordinate" definition of the Riemann tensor, and this proves that both definitions coincide.

5.5.4. ALGEBRAIC PROPERTIES OF THE RIEMANN CURVATURE TENSOR

Theorem 2. *For any three smooth fields X, Y, and Z on M^n the following identities are satisfied:*

(1) $R(X, Y)Z + R(Y, X)Z = 0$, $R^i_{j,kl} + R^i_{j,lk} = 0$ *(skew symmetry in the arguments X and Y).*

(2) $R(X, Y)Z + R(Z, X)Y + R(Y, Z)X = 0$ *(the Jacobi identity) or in the coordinate writing $R^i_{j,kl} + R^i_{l,jk} + R^i_{kl,j} = 0$. The correspondence between the indices (j, k, l) and fields X, Y, Z in $R^i_{j\,kl}$ is as follows: $j\sim X$, $k\sim Y$, $l\sim Z$.*

(3) *If the connection ∇ is Riemannian, we have $\langle R(X, Y)Z, W\rangle + \langle R(X, Y)W, Z\rangle = 0$ for arbitrary fields X, Y, Z, W, where $\langle,\rangle$ is the scalar product induced by the metric g_{ij}; in coordinate notation: $R_{ij,kl} + R_{ji,kl} = 0$, where $R_{ij,kl} = g_{i\alpha}R^\alpha_{j,kl}$.*

(4) *If the connection ∇ is Riemannian, $\langle R(X, Y)Z, W\rangle = \langle R(Z, W)X, Y\rangle$, i.e. $R_{ij,kl} = R_{kl,ij}$.*

In terms of the components $R_{ij,kl}$ (i.e. after index lowering) the Riemann tensor is skew-symmetric with respect to the indices in each pair (i, j) and (k, l), and is symmetric with respect to the per-

mutation of the pairs (when the indices in each pair remain in their own places).

Proof. (1) Evidently, the relation $R(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$ implies that $R(X, Y)Z$ is skew-symmetric with respect to the pair X, Y.

(2) We first prove that for symmetric connection the relation $\nabla_X Y - \nabla_Y X = [X, Y]$ is valid. Indeed, we have in terms of the coordinates

$$\begin{aligned}\nabla_X Y - \nabla_Y X = X^i\nabla_i Y - Y^i\nabla_i X &= \Big\{X^i\left(\frac{\partial Y^k}{\partial x^i} + Y^p\Gamma^k_{ip}\right) \\ &\quad - Y^i\left(\frac{\partial X^k}{\partial x^i} + X^p\Gamma^\kappa_{ip}\right)\Big\}\frac{\partial}{\partial x^k} \\ &= \left\{X^i\frac{\partial Y^k}{\partial x^i} - Y^i\frac{\partial X^k}{\partial x^i} + X^iY^p\Gamma^k_{ip} - Y^iX^p\Gamma^k_{ip}\right\}\frac{\partial}{\partial x^k} \\ &= \left\{X^i\frac{\partial Y^k}{\partial x^i} - Y^i\frac{\partial X^k}{\partial x^i}\right\}\frac{\partial}{\partial x^k} = [X, Y],\end{aligned}$$

because $Y^iX^p\Gamma^k_{ip} = Y^pX^i\Gamma^k_{pi}$, $\Gamma^k_{pi} = \Gamma^k_{ip}$. If X and Y are commutative, then $\nabla_X Y = \nabla_Y X$. Let us now prove the Jacobi identity. By virtue of Theorem 1, it is sufficient to verify this identity only for commutative fields X, Y, and Z (e.g. for ∂_i, ∂_j, ∂_k). It suffices to demonstrate that

$$\begin{aligned}&\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z + \nabla_Z\nabla_X Y - \nabla_X\nabla_Z Y \\ &\quad - \nabla_{[Z,X]}Y + \nabla_Y\nabla_Z X - \nabla_Z\nabla_Y X - \nabla_{[Y,Z]}X \equiv 0.\end{aligned}$$

The fields X, Y, and Z are commutative, so that the identity in question follows from the relations of the type $\nabla_X Y = \nabla_Y X$. Condition (2) is thus proved.

(3) It is required to prove that $\langle R(X, Y)Z, W\rangle + \langle R(X, Y)W, Z\rangle = 0$. It suffices to verify that $\langle R(X, Y)Z, Z\rangle = 0$ (taking into account polarization of quadratic forms). As before, we assume that $[X, Y] = 0$. Then, $\langle R(X, Y)Z, Z\rangle = \langle(\nabla_X\nabla_Y - \nabla_Y\nabla_X)Z, Z\rangle$. We need to show that $\langle\nabla_X\nabla_Y Z, Z\rangle = \langle\nabla_Y\nabla_X Z, Z\rangle$. Consider the function $\langle Z, Z\rangle = f$ and calculate $X(f) = X\langle Z, Z\rangle = \nabla_X\langle Z, Z\rangle = 2\langle\nabla_X Z, Z\rangle$. Furthermore,

$$YX(f) = 2\nabla_Y\langle\nabla_X Z, Z\rangle = 2\langle\nabla_Y\nabla_X Z, Z\rangle + 2\langle\nabla_X Z, \nabla_Y Z\rangle.$$

Similarly,

$$XY(f) = 2\langle\nabla_X\nabla_Y Z, Z\rangle + 2\langle\nabla_Y Z, \nabla_X Z\rangle.$$

According to the symmetry of $\langle,\rangle$, we have $\langle\nabla_Y\nabla_X Z, Z\rangle = \langle\nabla_X\nabla_Y Z, Z\rangle$, which is what was required.

(4) Let us consider the octahedron shown in Fig. 5.39. Four of its faces are shaded, and each vertex is marked with a scalar product. We assert that the sum of the products at the vertices of each shaded

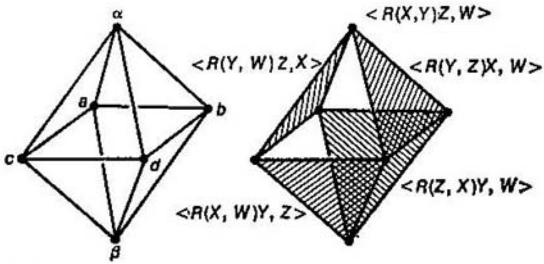

Figure 5.39

face is zero. Verify this for, say, face ($a\alpha c$). Using the symmetry relations which have already been proved, we have

$$\langle R(X, Y)Z, W\rangle = -\langle R(X, Y)W, Z\rangle,$$
$$\langle R(Y, W)Z, X\rangle = -\langle R(Y, W)X, Z\rangle,$$
$$\langle R(X, W)Y, Z\rangle = -\langle R(W, X)Y, Z\rangle,$$

i.e., according to the Jacobi identiy,

$$\begin{aligned} &a + \alpha + c \\ &= \langle R(X, Y)Z, W\rangle + \langle R(Y, W)Z, X\rangle + \langle R(X, W)Y, Z\rangle \\ &= -\langle R(X, Y)W, Z\rangle - \langle R(Y, W)X, Z\rangle - \langle R(W, X)Y, Z\rangle \\ &= -\langle R(X, Y)W + R(Y, W)X + R(W, X)Y, Z\rangle = 0. \end{aligned}$$

Similarly, we can verify that the sums $\alpha + b + d$, $c + d + \beta$, and $a + b + \beta$ vanish. Let us consider the identity $0 + 0 = 0 = 0 = 0 + 0$ and write out each of these zeros as follows: $(a + \alpha + c) + (\alpha + b + d) = 0 = (a + b + \beta) + (c + d + \beta)$, whence $2\alpha = 2\beta$, i.e. $\langle R(X, Y)Z, W\rangle = \langle R(Z, W)X, Y\rangle$, which completes the proof of the theorem.

Definition. The *Ricci tensor* of the Riemannian connection is the tensor $R_{jl} = R^i_{j,il}$, i.e. the tensor obtained by contracting the Riemann tensor with respect to a pair of indices. The Ricci tensor is symmetric (verify!).

Definition. The *scalar curvature* R of a Riemannian manifold is the function $R(x) = g^{kl}R_{kl}$, i.e. complete contraction of the Ricci tensor with the tensor inverse to the metric one.

Apparently, R_{kl} is a second-rank tensor, and $R(x)$ is a scalar function. For many particular problems it is useful to express a Riemann tensor explicitly in terms of g_{ij} and its derivatives.

Theorem 3. *On a Riemannian manifold the following identity holds true*:

$$\begin{aligned} R_{iq,kl} &= g_{i\alpha}R^{\alpha}_{q,kl} \\ &= \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial x^q\,\partial x^k}+\frac{\partial^2 g_{qk}}{\partial x^i\,\partial x^l}-\frac{\partial^2 g_{ik}}{\partial x^q\,\partial x^l}-\frac{\partial^2 g_{ql}}{\partial x^i\,\partial x^k}\right) \\ &+ g_{mp}(\Gamma^m_{qk}\Gamma^p - \Gamma^m_{ql}\Gamma^p_{ik}). \end{aligned}$$

Proof. The coordinate representation of the Riemann tensor implies that

$$\begin{aligned} R^i_{q,kl} &= \frac{\partial \Gamma^i_{ql}}{\partial x^k} - \frac{\partial \Gamma^i_{qk}}{\partial x^l} + \Gamma^p_{ql}\Gamma^i_{pk} - \Gamma^p_{qk}\Gamma^i_{pl} \\ &= \left(\frac{\partial \Gamma^i_{ql}}{\partial x^k} + \Gamma^p_{ql}\Gamma^i_{pk}\right)\cdot[k, l], \end{aligned}$$

where $[k, l]$ denotes alternation with respect to indices k and l without division by 2. Furthermore,

$$\begin{aligned} \theta = g_{si}R^i_{q,kl} = R_{sq,kl} &= g_{si}\left(\frac{\partial \Gamma^i_{ql}}{\partial x^k} + \Gamma^p_{ql}\Gamma^i_{pk}\right)[k, l] \\ &= g_{si}\left(\frac{\partial \Gamma^i_{*}}{\partial x^k} + \Gamma^i_{pk}\Gamma^p_{*}\right)[k, l], \end{aligned}$$

where $*$ stands for the pair of indices (ql). The expression in parentheses may, formally, be treated as the result of covariant differentiation ∇_k of the set $\{\Gamma^i_*\}$ (the symbol $*$ is ignored for the time being). The set $\{\Gamma^i_*\}$ does not form a tensor, but in each given coordinate system we may also consider a tensor with the components Γ^i_* (in other systems this tensor will have components distinct from Γ^i_*, although this fact is insignificant for differentiation in a given system). Since g_{si} induces index lowering in the "tensor" Γ^i_*, this operation can be performed under the sign of covariant differentiation because the tensor g_{ij} is covariantly constant. It follows that

$$\theta = g_{si}\nabla_k(\Gamma^i_*)[k, l] = \nabla_k(g_{si}\Gamma^i_*)[k, l] = \nabla_k(\Gamma_{s,*})[k, l],$$

where

$$\begin{aligned} \Gamma_{s,*} = \Gamma_{s,ql} &= \frac{1}{2}g_{si}g^{i\alpha}\left(\frac{\partial g_{\alpha l}}{\partial x^q} + \frac{\partial g_{\alpha q}}{\partial x^l} - \frac{\partial g_{ql}}{\partial x^\alpha}\right) \\ &= \frac{1}{2}\left(\frac{\partial g_{sl}}{\partial x^q} + \frac{\partial g_{sq}}{\partial x^l} - \frac{\partial g_{ql}}{\partial x^s}\right). \end{aligned}$$

Substitution into the initial formula for $R_{sq,kl}$ yields

$$R_{sq,kl}=\nabla_k(\Gamma_{s_*})[k,\ l]=\left(\frac{\partial\Gamma_{s*}}{\partial x^k}-\Gamma_{\alpha,_*}\Gamma^{\alpha}_{ks}\right)[k,\ l]$$

$$=\left(\frac{\partial\Gamma_{s,ql}}{\partial x^k}-\Gamma_{\alpha,ql}\Gamma^{\alpha}_{ks}\right)[k,\ l]$$

$$=\frac{1}{2}\left(\frac{\partial^2 g_{sl}}{\partial x^k\,\partial x^q}+\frac{\partial^2 g_{sq}}{\partial x^k\,\partial x^l}-\frac{\partial^2 g_{ql}}{\partial x^k\,\partial x_s}\right)[k,\ l]$$

$$-g_{\alpha p}\Gamma^p_{ql}\Gamma^{\alpha}_{ks}[k,\ l]$$

$$=\frac{1}{2}\left(\frac{\partial^2 g_{sl}}{\partial x^k\,\partial x^q}+\frac{\partial^2 g_{qk}}{\partial x^l\,\partial x^s}-\frac{\partial^2 g_{sk}}{\partial x^l\,\partial x^q}-\frac{\partial^2 g_{ql}}{\partial x^k\,\partial x^l}\right)$$

$$+g_{\alpha p}(\Gamma^p_{qk}\Gamma^{\alpha}_{sl}-\Gamma^p_{ql}\Gamma^{\alpha}_{ks}).$$

The theorem is proved.

Corollary 1. *If the Riemann curvature tensor does not vanish in a certain coordinate system, no local Euclidean coordinates can be defined on M^n, i.e. the coordinates in which g_{ij} is a constant matrix (or, which is the same, $\Gamma^i_{jk}\equiv 0$).*

The *proof* follows from Lemma 2.

We can also reason as follows. Let us consider the transformation law $\Gamma^{i'}_{j'k'}=\frac{\partial x^{i'}}{\partial x^i}\left(\frac{\partial x^j}{\partial x^{j'}}\Gamma^i_{jk}+\frac{\partial^2 x^i}{\partial x^{j'}\,\partial x^{k'}}\right)$. For the coordinates, in which $\Gamma^{i'}_{j'k'}\equiv 0$, to exist it is necessary that the following relations (i.e. equations for the coordinates $x^{i'}$) be valid:

$$\frac{\partial^2 x^i}{\partial x^{j'}\,\partial x^{k'}}=-\frac{\partial x^j\,\partial x^k}{\partial x^{j'}\,\partial x^{k'}}\Gamma^i_{jk}.$$

The necessary condition for the solvability of this system is given by the identities $\frac{\partial}{\partial x^{\alpha'}}\left(\frac{\partial^2 x^i}{\partial x^{j'}\,\partial x^{k'}}\right)=\frac{\partial}{\partial x^{k'}}\left(\frac{\partial^2 x^i}{\partial x^{j'}\,\partial x^{\alpha'}}\right)$, and this in turn imposes the conditions on the right-hand sides of the system. It can be proved that the fulfilment of these conditions is equivalent to the vanishing of the curvature tensor (verify!).

5.5.5. CERTAIN APPLICATIONS OF THE RIEMANN CURVATURE TENSOR

For a two-dimensional Riemannian manifold the curvature tensor is especially simple (problem: what is the meaning of the curvature tensor for a one-dimensional manifold?). Let us consider the scalar curvature $R(x)$, a function on M^2. Since this function is a measure of "curvature" of M^2, there is every reason to believe that

it is related to the Gaussian curvature which is also known to be responsible for the "curvature" of M^2.

Theorem 4. *On a two-dimensional smooth Riemannian manifold the identity $R = 2K$ holds true, where $R(P)$, $P \in M^2$, is the scalar curvature and $K(P)$ is the Gaussian curvature.*

Corollary 2. *Since $R(P)$ is completely defined by g_{ji}, $K(P)$ is also completely defined by g_{ij}; in particular, $K(P)$ does not change under isometries of M^2 in $\mathbf{R}^3$ (i.e. under surface bending).*

This corollary is not trivial, for the definition of $K(P)$ relies upon the second fundamental form which describes the embedding of M^2 in $\mathbf{R}^3$. A straightforward verification of the invariance of $K(P)$ under bending is not simple either and can be performed most easily only after studying the properties of the Riemann tensor.

Proof of the theorem. By virtue of Theorem 3,

$$R_{iq,kl} = \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial x^q\,\partial x^k} + \frac{\partial^2 g_{qk}}{\partial x^i\,\partial x^l} - \frac{\partial^2 g_{ik}}{\partial x^q\,\partial x^l} - \frac{\partial^2 g_{ql}}{\partial x^i\,\partial x^k}\right) + g_{mp}\left(\Gamma^m_{qk}\Gamma^p_{il} - \Gamma^m_{ql}\Gamma^p_{ik}\right).$$

Let us introduce in $\mathbf{R}^3$ a special Cartesian coordinate system, choose on M^2 a point P, and define M^2 in the neighbourhood of P as the graph $z = f(x, y)$, where (x, y) are Cartesian coordinates in T_PM^2. Since $T_PM^2 = \mathbf{R}^2(x, y)$ is the tangent plane, $\operatorname{grad} f(P) = 0$, i.e. $g_{ij}(P) = (\delta_{ij} + f_{x^i}f_{x^j})(P) = \delta_{ij}$ or $\Gamma^i_{jk}(P) = 0$ because $(\partial g_{ij}/\partial x^k)|_P = 0$ (verify!). As the tensor $R_{iq,kl}$ is algebraically symmetric, it has only one principal component, $R_{12,12}$. The others either vanish, or differ from $R_{12,12}$ only in sign, or coincide with $R_{12,12}$. Writing the Riemann tensor in the coordinate system (x, y), we obtain

$$\begin{aligned}
R_{12,12} &= \frac{1}{2}\left(2\cdot\frac{\partial^2 g_{12}}{\partial x\,\partial y} - \frac{\partial^2 g_{22}}{\partial x^2} - \frac{\partial^2 g_{11}}{\partial y^2}\right)\\
&= \frac{1}{2}\left[2(f_xf_y)_{xy} - (f_y^2)_{xx} - (f_x^2)_{yy}\right]\\
&= (f_{xx}f_y + f_xf_{xy})_y - (f_yf_{xy})_x - (f_xf_{xy})_y\\
&= f_{xxy}f_y + f_{xx}f_{yy} + f_{xy}f_{xy} + f_xf_{xyy}\\
&\quad - f_{xy}f_{xy} - f_yf_{xxy} - f_{xy}f_{xy} - f_xf_{xyy}\\
&= \begin{vmatrix} f_{xx} & f_{xy}\\ f_{xy} & f_{yy}\end{vmatrix} = K,
\end{aligned}$$

whence $R_{12,12} = K$. Calculation of R yields

$$R = g^{kl}R_{kl} = g^{kl}R^{\alpha}_{k,\alpha l} = g^{kl}g^{\alpha q}R_{qk,\alpha l} = R_{12,12}\left(\Sigma \pm g^{kl}\cdot g^{\alpha q}\right),$$

where

$$\begin{aligned}\Sigma \pm g^{kl}g^{\alpha q} &= g^{22}g^{11} - g^{21}g^{21} + g^{11}g^{22} - g^{12}g^{12} \\ &= 2\left(g^{22}g^{11} - (g^{12})^2\right) = 2\det(g_{ij})^{-1} \\ &= \frac{2}{g}, \text{ with } g = \det g_{ij}.\end{aligned}$$

Thus, $R = \frac{2}{g} R_{12,12}$, so that $R = 2K$ because $g_{ij}(P) = \delta_{ij}$. But R and K are scalars and their values do not depend on the choice of the coordinate system, hence, in any system $R = 2K$.

We have demonstrated that the behaviour of $K(P)$ under bending differs from the corresponding behaviour of the mean curvature, and therefore the Gaussian curvature is an "intrinsic invariant" of a surface. Let us consider several examples.

(1) For the Euclidean metric $dx^2 + dy^2$, $R = 2K = 0$,

(2) for the spherical metric $dr^2 + \left(\sin^2 \frac{r}{r_0}\right) d\varphi^2$, $R = 2K = \frac{2}{r_0^2}$, i.e. the scalar curvature is constant and positive,

(3) for the metric of the Lobachevskian plane $dr^2 + \left(\sinh \frac{r}{r_0}\right)^2 \cdot d\varphi^2$ we have $R = 2K = -\frac{2}{r_0^2}$, i.e. the curvature is constant and negative,

(4) for the conformal Euclidean metric $\lambda(x, y)(dx^2 + dy^2)$, where $\lambda(x, y)$ is a positive function, $R = 2K = -\frac{1}{\lambda}\Delta \ln \lambda$, where Δ is the Laplacian. This formula can be proved by straightforward calculation.

In three dimensions the Riemann tensor has a more complicated structure. The number of the principal components increases up to six: $R_{12,13}$, $R_{21,23}$, $R_{31,32}$, $R_{12,12}$, $R_{13,13}$, and $R_{23,23}$; the other components $R_{ij,kl}$ either vanish or coincide with the principal components, or differ from them only in sign.

The "intricacy" of $R_{ij,kl}$ depends on the number of principal components. There is only one such component for M^2 and six components for M^3. It can be shown that for a manifold of an arbitrary dimension n the number of principal components is $N = \frac{n^2(n^2-1)}{12}$, for $n \to \infty$ the ratio of N to the total number of components (i.e. to n^4) tends to 1/12 (problem: find N).

Of great importance in geometry is "curvature along a two-dimensional direction". Let us consider a Riemannian manifold M^n and let $X, Y \in T_P M^n$. Suppose these vectors are so chosen that the area of the parallelogram $\Pi(X, Y)$ constructed on these vectors is unity in the metric g_{ij}. Then, the *curvature of M^n along the two-dimensional*

direction σ defined by **X**, **Y** is the number $R(\sigma) = \langle R(X, Y)X, Y\rangle$ where X and Y are arbitrary vector fields in the neighbourhood of P such that $X(P) = \mathbf{X}$, $Y(P) = \mathbf{Y}$, i.e. they coincide at P with the vectors **X**, **Y**. It is required to prove that $R(\sigma)$ does not depend on the way **X**, **Y** are included in the vector fields X, Y.

Lemma 3. *There is valid the formula* $R(\sigma) = R_{\beta j,kl} X^j X^k Y^l Y^\beta$, *where* X^α, Y^β *are the coordinates of the vectors* **X**, **Y**. *Here* $R_{ij,kl}$ *is a tensor, and* $R(\sigma)$ *does not depend on the way* **X**, **Y** *are included in* X, Y.

Proof. We have

$$[R(X, Y)Z]^k = R^k_{j,pq} X^p Y^q Z^j,$$

$$R(\sigma) = g_{\alpha\beta} Y^\beta [R(X, Y)Y]^\alpha = g_{\alpha\beta} Y^\beta R^\alpha_{j,kl} X^j X^k Y^l = R_{\beta j,kl} X^j X^k Y^l Y^\beta,$$

which is what was required because $\mathbf{X}^i = X^i(P)$ and $\mathbf{Y}^i = Y^i(P)$.

Definition. A Riemannian manifold M^n is called a *manifold of positive (constant, negative, zero, etc.) curvature* if its curvatures along all two-dimensional directions are positive (constant, negative, zero, etc.).

To justify this definition, we should compare it with the "two-dimensional" definitions of the positive, etc., curvatures in terms of the Gaussian curvature.

Lemma 4. *Given a Riemannian manifold* M^2, *and let* $K(P)$ *be the Gaussian curvature*, $R(P)$ *the scalar curvature, and* $R(\sigma)$ *the curvature along a two-dimensional direction* $(\mathbf{X}, \mathbf{Y}) = \sigma$ *at a point* $P \in M^2$. *Then* $R(\sigma) = K(P) = \frac{1}{2} R(P)$.

Proof. Let us consider the coordinates (x, y) in which coordinate lines are orthogonal at the point P. We may assume the scalar product induced by the metric on $T_P M^2$ to be Euclidean. Then,

$$\begin{aligned} R(\sigma) &= R_{\beta j,kl} X^j X^k Y^\beta Y^l \\ &= R_{12,12}(X^2X^2Y^1Y^1 - X^2X^1Y^2Y^1 + X^1X^1Y^2Y^2 - X^1X^2Y^1Y^2) \\ &= R_{12,12}(X^2Y^1 - Y^2X^1)^2 = R_{12,12}\cdot 1 = R_{12,12}, \end{aligned}$$

because $X^2Y^1 - Y^2X^1 =$ (the area of the parallelogram $\Pi(X, Y)$ spanned by **X**, **Y** in the Cartesian coordinates (x, y) in $T_P(M^2)$. Since $R_{12,12} = K(P)$, we obtain $R(\sigma) = K(P)$, which is what was required.

Thus, the definition of the curvature along a two-dimensional direction suggested above is a natural generalization of the concept of the Gaussian (scalar) two-dimensional curvature of the Riemannian metric. Evidently, the curvatures along any two-dimensional direction in $\mathbf{R}^n$ vanish, i.e. $\mathbf{R}^n$ is a manifold of zero curvature.

The curvature along a two-dimensional direction $\sigma = (\mathbf{X}, \mathbf{Y})$ admits a clear interpretation, which is presented here without proof. Let us consider at a point $P \in M^n$ a two-dimensional plane H spanned by $\mathbf{X}$, $\mathbf{Y}$ and draw from P geodesics $\gamma_{\mathbf{Z}}(t)$ along each vector

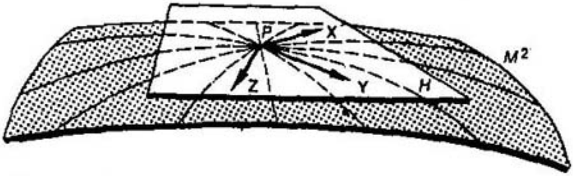

Figure 5.40

$\mathbf{Z} \in H$. Apparently, these geodesics form (locally) a two-dimensional surface $M^2 \subset M^n$ such that the tangent plane to it at P coincides with H (see Fig. 5.40). This surface is called a *geodesic surface*. On a geodesic surface there exists an induced metric which admits definition of the Gaussian curvature; it appears that $\mathbf{R}(\sigma)$ coincides with this Gaussian curvature.

Chapter 6

Homology Theory

Thus far, our interest has mainly been focused on the local properties of a smooth manifold, i.e. the properties which can be studied independently in the neighbourhood of each point P of a manifold M and do not depend on the way M is defined as a union of charts. In many problems, however, it is insufficient to know only the local properties of M.

Let us consider, for example, the following problem. Let $M = S^1$ be a circle with the angular parameter φ chosen as a local coordinate. We now try to find a smooth function f on S^1 such that the identity

$$\frac{df}{d\varphi} = g(\varphi) \tag{1}$$

is satisfied, where g is a smooth function on S^1. If we solve this problem in a small neighbourhood of a point $P \in S^1$ with the coordinate φ_0, any antiderivative $f(\varphi) = \int g(\varphi)\, d\varphi$ of the function g may be a solution. Generally, the problem does not always has a solution on S^1. Indeed, any smooth function on S^1 can be identified with a 2π-periodic function of one real variable. Then the function f is a solution of our problem, provided $f(\varphi) = \int g(\varphi)\, d\varphi$ and f is periodic. Any antiderivative of the function g can be expressed in terms of the definite integral $f(\varphi) = \int\limits_{\varphi_0}^{\varphi} g(\varphi)\, d\varphi + C$. The condition that f is periodic can, therefore, be written in the form

$$\int\limits_{\varphi_0}^{\varphi_0+2\pi} g(\varphi)\, d\varphi = 0 \quad \text{or} \quad \int\limits_{0}^{2\pi} g(\varphi)\, d\varphi = 0. \tag{2}$$

For instance, if $g(\varphi) \equiv 1$, then $\int\limits_0^{2\pi} g(\varphi)\,d\varphi = 2\pi \neq 0$. Thus, for $g(\varphi) \equiv 1$ problem (1) does not have a solution on S^1. Condition (2) is the necessary and sufficient condition for the existence of a solution of problem (1). Here we see that the existence of a solution depends essentially on the structure of a manifold as a whole. Section 6.1 deals with those properties of a manifold that govern the global behaviour of a function or a mapping on manifold.

6.1. CALCULUS OF EXTERIOR DIFFERENTIAL FORMS. COHOMOLOGY GROUPS

6.1.1. DIFFERENTIATION OF EXTERIOR DIFFERENTIAL FORMS

Differential calculus of exterior differential forms is somewhat simpler than differential calculus of arbitrary tensor fields. If a connection is valid on a manifold, the covariant gradient of a skew-symmetric tensor field is not, in general, a skew-symmetric tensor field. It is natural, therefore, to define the gradient of an exterior differential form as the composite of gradient covariant component and alternation with respect to all indices of the tensor field. If the connection is symmetric, the Christoffel symbols do not appear in the formula for the gradient of an exterior differential form, i.e. the definition of this gradient does not depend on the choice of symmetric connection on the manifold. Indeed, let in the local coordinate system $(x^1, \ldots, x^n)$ the differential form ω have the components $\{\omega_{i_1}, \ldots, \omega_{i_k}\}$. Then the gradient $d\omega$ is defined by

$$(d\omega)_{j_1, \ldots, j_{k+1}} = \sum_\sigma (-1)^{|\sigma|} \nabla_{\sigma(j_{k+1})} \omega_{\sigma(j_1), \ldots, \sigma(j_k)}, \tag{1}$$

where

$$\nabla_l \omega_{i_1, \ldots, i_k} = \frac{\partial \omega_{i_1, \ldots, i_k}}{\partial x^l} - \Gamma^\alpha_{li_1} \omega_{\alpha i_2, \ldots, i_k} - \ldots - \Gamma^\alpha_{li_k} \omega_{i_1, \ldots, \alpha}. \tag{2}$$

Substitution of (2) into (1) yields

$$(d\omega)_{j_1, \ldots, j_{k+1}} = \sum_\sigma (-1)^{|\sigma|} \frac{\partial \omega_{\sigma(j_1), \ldots, \sigma(j_k)}}{\partial x^{\sigma(j_{k+1})}}$$

$$- \sum_\sigma \sum_{s=1}^{k} (-1)^{|\sigma|} \Gamma^\alpha_{\sigma(j_{k+1})\sigma(j_s)} \omega_{\sigma(j_1) \ldots \sigma(j_{s-1})\sigma(j_{s+1}) \ldots \sigma(j_k)}.$$

The second term vanishes because for fixed s and α there exist exactly two permutations of indices $(j_1, \ldots, j_{k+1})$, σ and σ', such that $\sigma(j_1) = \sigma'(j_1), \ldots, \sigma(j_{s-1}) = \sigma'(j_{s-1})$ and $\sigma(j_{s+1}) = \sigma'(j_{s+1}), \ldots, \sigma(j_k) = \sigma'(j_k)$, and also $\sigma(j_{k+1}) = \sigma'(j_s)$, $\sigma(j_s) = \sigma'(j_{k+1})$; the parity of these permutations is distinct. Thus, since the Christoffel symbols Γ^i_{jk} are symmetric in the lower indices, the two terms corresponding to the permutations σ and σ' are cancelled out, so that we obtain

$$(d\omega)_{j_1, \ldots, j_{k+1}} = \sum_{\sigma} (-1)^{|\sigma|} \frac{\partial \omega_{\sigma(j_1), \ldots, \sigma(j_k)}}{\partial x^{\sigma(j_{k+1})}}. \tag{3}$$

In formula (3) many of the terms appear several times. Indeed, if $\sigma(j_{k+1}) = \sigma'(j_{k+1})$, then $\sigma = \sigma' h$, where h is the permutation which leaves j_{k+1} in its own place. In this case

$$(-1)^{|\sigma'|} \frac{\partial \omega_{\sigma'(j_1) \ldots \sigma'(j_k)}}{\partial x^{\sigma'(j_{k+1})}} = (-1)^{|\sigma|+|h|} \frac{\partial \omega_{\sigma h(j_1) \ldots \sigma h(j_k)}}{\partial x^{\sigma(j_{k+1})}}$$

$$= (-1)^{|\sigma|} \frac{\partial \omega_{\sigma(j_1) \ldots \sigma(j_k)}}{\partial x^{\sigma(j_{k+1})}}.$$

Bearing in mind what has just been said (and trying to make gradient and exterior product compatible operations), we retain in the components of the gradient $d\omega$ only one term out of identical $k!$ terms. We arrive therefore at the following definition.

Definition 1. The *gradient of an exterior differential form* ω is an exterior differential form $d\omega$ whose components in a local coordinate system $(x^1, \ldots, x^n)$ are written as

$$(d\omega)_{j_1, \ldots, j_{k+1}} = \sum_{s=1}^{k+1} (-1)^{s+1} \frac{\partial \omega_{j_1 \ldots j_{s-1} j_{s+1} \ldots j_{k+1}}}{\partial x^{j_s}}. \tag{4}$$

Formula (4) differs from formula (3) by the factor and sign.

Theorem 1. *The gradient of an exterior differential form satisfies the following conditions*: (a) $d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^{\deg \omega_1} \omega_1 \wedge d\omega_2$, (b) $d(d\omega) = 0$.

Proof. Let $(x^1, \ldots, x^n)$ be a local coordinate system and let ω_1, ω_2 be two exterior differential forms of degrees p and q, respectively. Then the components of ω_1 and ω_2 do not vanish only for those sets of indices $(i_1, \ldots, i_p)$ and $(j_1, \ldots, j_q)$ for which $i_s \neq i_l$, $j_s \neq j_l$ for $s \neq l$. If $i_1 < i_2 < \ldots < i_p$, $I = \{i_1, \ldots, i_p\}$ and $j_1 < j_2 < \ldots < j_q$, $J = \{j_1, \ldots, j_q\}$, we put $\omega_{1,I} = \omega_{1, i_1, \ldots, i_p}$, $\omega_{2,J} = \omega_{2, j_1, \ldots, j_q}$. Similarly, if $l_1 < \ldots < l_{p+q}$, $K = \{l_1, \ldots, l_{p+q}\}$,

we put $(\omega_1 \wedge \omega_2)_K = (\omega_1 \wedge \omega_2)_{i_1, \ldots, i_{p+q}}$. By formula (1) of Sec. 5.2, we obtain

$$(\omega_1 \wedge \omega_2)_K = \sum (-1)^{|\sigma(I)|} \omega_{1,I}\omega_{2,J}, \tag{5}$$

where summation extends over all the subsets $I \subset K$ consisting of p elements, $I = K \setminus J$, and σ is the permutation which places the indices of the set I into the first p positions retaining their order. Formula (4) for the gradient can also be rewritten as

$$(d\omega)_I = \sum_{i \in I} (-1)^{|\sigma(i)|} \frac{\partial \omega_{I \setminus \{i\}}}{\partial x^i}, \tag{6}$$

where $\sigma(i)$ is the permutation of the set I which places index i into the first position, retaining the order of the remaining indices. Applying (6) to (5), we obtain

$$\begin{aligned}
&(d(\omega_1 \wedge \omega_2))_K \\
&\quad = \sum_{i \in K} (-1)^{|\sigma(i)|} \frac{\partial (\omega_1 \wedge \omega_2)_{K > \{i\}}}{\partial x^i} \\
&\quad = \sum_{i \in K} (-1)^{|\sigma(i)|} \sum_{I} (-1)^{|\sigma(I)|} \frac{\partial (\omega_{1,I} \cdot \omega_{2,K \setminus I \setminus \{i\}})}{\partial x^i} \\
&\quad = \sum_{K = I \cup J \cup \{i\}} (-1)^{|\sigma(I,J,i)|} \frac{\partial (\omega_{1,I} \cdot \omega_{2,J})}{\partial x^i}.
\end{aligned}$$

In the last sum summation extends over all possible partitions of the set K into three disjoint subsets $\{i\}$, I, J, and $\sigma(I, J, i)$ is the permutation which puts index i into the first place and interchanges the sets I and J retaining the order of indices within them. Thus,

$$\begin{aligned}
&(d(\omega_1 \wedge \omega_2))_K \\
&\quad = \sum_{K = I \cup J \cup \{i\}} (-1)^{|\sigma(i,I,J)|} \frac{\partial (\omega_{1,I} \cdot \omega_{2,J})}{\partial x^i} \\
&\quad = \sum_{K = I \cup J \cup \{i\}} (-1)^{|\sigma(i,I,J)|} \frac{\partial \omega_{1,I}}{\partial x^i} \omega_{2,J} \\
&\qquad + \sum_{K = I \cup J \cup \{i\}} (-1)^{|\sigma(i,I,J)|} \omega_{1,I} \frac{\partial \omega_{2,J}}{\partial x^i} \\
&\quad = (d\omega_1 \wedge \omega_2)_K + (-1)^{\deg \omega_1} (\omega_1 \wedge d\omega_2)_K.
\end{aligned}$$

It is sufficient to prove property (b) (in Theorem 1) for the basic forms into which any form can be decomposed. We may assume, without loss of generality, that in the local coordinate system

$(x^1, \ldots, x^n)$ the form ω can be written as $\omega = f(x^1, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^k$. If condition (b) is valid for the forms ω_1 and ω_2, it is also valid for their exterior product $\omega_1 \wedge \omega_2$. Indeed,

$$\begin{aligned} dd(\omega_1 \wedge \omega_2) &= d\left(d\omega_1 \wedge \omega_2 + (-1)^{\deg \omega_1} \omega_1 \wedge d\omega_2\right) \\ &= dd\omega_1 \wedge \omega_2 + (-1)^{\deg (d\omega_1)} d\omega_1 \wedge d\omega_2 \\ &\quad + (-1)^{\deg \omega_1} d\omega_1 \wedge d\omega_2 + (-1)^{2\deg \omega_1} \omega_1 \wedge dd\omega_2. \end{aligned}$$

The first and last terms are zero by assumption, and the other two are equal but have opposite signs. It suffices, therefore, to verify condition (b) for an arbitrary function f and for the form dx^k. In the latter case $dd\,(dx^k) = 0$, since x^k is also a function on the manifold. We now demonstrate that $d\,(df) \equiv 0$. Indeed, $(df)_i = \frac{\partial f}{\partial x^i}$ and also

$$(d\,(df))_{ij} = \frac{\partial\,(df)_j}{\partial x^i} - \frac{\partial\,(df)_i}{\partial x^j} = \frac{\partial^2 f}{\partial x^i\, \partial x^j} - \frac{\partial^2 f}{\partial x^j\, \partial x^i} \equiv 0.$$

Theorem 1 is proved.

It should be noted that condition (b) relies upon the remarkable property of a function of many variables: the independence of mixed partial derivatives of differentiation order. Condition (b) is, in certain senses, an ultimate extension of this property to systems of functions that form the components of a skew-symmetric tensor.

Examples

1. In a local coordinate system any exterior differential form can, by definition, be represented as the sum $\omega = \sum\limits_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p}$. Thus, according to Theorem 1, the gradient of the form ω is calculated by the formula

$$d\omega = \sum_{i_1 < \ldots < i_p} d\,(\omega_{i_1 \ldots i_p}) \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}. \tag{7}$$

Furthermore, in a local coordinate system the gradient of a smooth function, df, has the components $\left\{\frac{\partial f}{\partial x^i}\right\}$. Hence, $df = \Sigma \frac{\partial f}{\partial x^i}\, dx^i$ and therefore

$$d\omega = \sum_{i_1 < \ldots < i_p} \sum_{i=1}^{n} \frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^i}\, dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

This formula is, apparently, equivalent to formula (4).

2. Let us find a differential form ω which satisfies the identity

$$d\omega = \Omega, \tag{8}$$

where Ω is a fixed differential form. In each local coordinate system equation (8) is reduced to the system of partial differential equations

$$\sum_{s=1}^{k+1} (-1)^{s+1} \frac{\partial \omega_{i_1 \ldots i_{s-1} i_{s+1} \ldots i_{k+1}}}{\partial x^{i_s}} = \Omega_{i_1 \ldots i_{k+1}}.$$

According to condition (b) of Theorem 1, equation (8) has solutions if and only if $d\Omega = 0$. Indeed, applying the operator d to the left-hand and right-hand sides of (8), we obtain $d\Omega = d\,(d\omega) = 0$. In particular, for $\deg \Omega = 1$ the problem is reduced to finding on a manifold M a smooth function f such that its gradient is equal to a given form Ω. In a local coordinate system $(x^1, \ldots, x^n)$ the form Ω is represented as $\Omega = \sum_{i=1}^{n} \Omega_i \, dx^i$, and equation (8) is equivalent to the system

$$\frac{\partial f}{\partial x^i} = \Omega_i, \quad i = 1, \ldots, n. \tag{9}$$

If system (9) has a solution, then, by differentiating (9) with respect to the variables x^j, we obtain $\frac{\partial^2 f}{\partial x^i \, \partial x^j} = \frac{\partial \Omega_i}{\partial x^j}$. Since mixed partial derivatives of a smooth function f do not depend on differentiation order, we have

$$\frac{\partial \Omega_i}{\partial x^j} = \frac{\partial \Omega_j}{\partial x^i}, \quad 1 \leqslant i, \quad j \leqslant n. \tag{10}$$

Conditions (10), when written without any reference to coordinates, are equivalent to $d\Omega = 0$. We now prove that conditions (10) are sufficient for system (9) to have a solution in a rather small neighbourhood of any point P of the manifold M. To this end, we first solve one of the equations of (9), say, $\frac{\partial f}{\partial x^1} = \Omega_1$. We have

$$f(x^1, \ldots, x^n) = \int_{x_0^1}^{x^1} \Omega_1 (x^1, x^2, \ldots, x^n) \, dx^1 + \varphi_2 (x^2, \ldots, x^n), \tag{11}$$

where $\varphi_2 (x^2, \ldots, x^n)$ is an arbitrary smooth function of the variables $(x^2, \ldots, x^n)$. Substitution of the right-hand side of (11) into the

second equation of (9) yields

$$\frac{\partial f}{\partial x^2} = \int_{x_0^1}^{x^1} \frac{\partial \Omega_1}{\partial x^2}(x^1, x^2, \ldots, x^n)\,dx^1 + \frac{\partial \varphi_2}{\partial x^2}(x^2, \ldots, x^n)$$
$$= \Omega_2(x^1, \ldots, x^n).$$

Taking into account (10), we obtain

$$\Omega_2(x^1, \ldots, x^n)$$
$$= \int_{x_0^1}^{x^1} \frac{\partial \Omega_2}{\partial x^1}(x^1, \ldots, x^n)\,dx^1 + \frac{\partial \varphi_2}{\partial x^2}(x^2, \ldots, x^n)$$
$$= \Omega_2(x^1, \ldots, x^n) - \Omega_2(x_0^1, x^2, \ldots, x^n) + \frac{\partial \varphi_2}{\partial x^2}(x^2, \ldots, x^n),$$

i.e.

$$\frac{\partial \varphi_2}{\partial x^2}(x^2, \ldots, x^n) = \Omega_2(x_0^1, x^2, \ldots, x^n). \tag{12}$$

The general solution of equation (12) is of the form

$$\varphi_2(x^2, \ldots, x^n) = \int_{x_0^2}^{x^2} \Omega_2(x_0^1, x^2, \ldots, x^n)\,dx^2 + \varphi_3(x^3, \ldots, x^n). \tag{13}$$

Substitution of (11) and (13) into the third equation of (9) yields

$$\Omega_3(x^1, \ldots, x^n) = \int_{x_0^1}^{x^1} \frac{\partial \Omega_1}{\partial x^3}(x^1, \ldots, x^n)\,dx^1$$
$$+ \int_{x_0^2}^{x^2} \frac{\partial \Omega_2}{\partial x^3}(x_0^1, x^2, \ldots, x^n)\,dx^2 + \frac{\partial \varphi_3}{\partial x^3}(x^3, \ldots, x^n).$$

Again, taking into account condition (10), we obtain the equation for the function φ_3

$$\Omega_3(x^1, \ldots, x^n)$$
$$= \int_{x_0^1}^{x^1} \frac{\partial \Omega_3}{\partial x^1}(x^1, \ldots, x^n)\,dx^1$$
$$+ \int_{x_0^2}^{x^2} \frac{\partial \Omega_3}{\partial x^2}(x_0^1, x^2, \ldots, x^n)\,dx^2 + \frac{\partial \varphi_3}{\partial x^3}(x^3, \ldots, x^n)$$
$$= \Omega_3(x^1, \ldots, x^n) - \Omega_3(x_0^1, x^2, \ldots, x^n) + \Omega_3(x_0^1, x^2, \ldots, x^n)$$
$$- \Omega_3(x_0^1, x_0^2, x^3, \ldots, x^n) + \frac{\partial \varphi_3}{\partial x^3}(x^3, \ldots, x^n),$$

$$\frac{\partial \varphi_3}{\partial x^3}(x^3, \ldots, x^n) = \Omega_3(x_0^1, x_0^2, x^3, \ldots, x^n).$$

Repeating the procedure, we arrive at the sequence of functions $\varphi_k(x^k, \ldots, x^n)$ defined by the following recurrence relations:

$$\varphi_k(x^k, \ldots, x^n) = \int_{x_0^k}^{x^k} \Omega_k(x_0^1, \ldots, x_0^{k-1}, x^k, \ldots, x^n)\, dx^k + \varphi_{k+1}(x^{k+1}, \ldots, x^n),$$

$$\varphi_{n+1} \equiv \text{const},$$

$$f(x^1, \ldots, x^n) = \sum_{k=1}^{n} \int_{x_0^k}^{x^k} \Omega_k(x_0^1, \ldots, x_0^{k-1}, x^k, \ldots, x^n)\, dx^k + \varphi_{n+1}.$$

The last relation is an arbitrary solution of system (9), which depends on one numerical parameter $\varphi_{n+1} = \text{const}$. Incidentally, nonuniqueness of the solution of system (9) follows from general algebraic properties of an exterior differential form. Indeed, if f_1 and f_2 are two solutions of system (9), then $df_1 = \Omega$ and $df_2 = \Omega$, i.e. $d(f_1 - f_2) = \Omega - \Omega = 0$. The function $h = f_1 - f_2$ has the property that its gradient is identically zero on a manifold M. Hence, h is a locally constant function. If M is connected, then h is a constant function. Thus, *any two solutions of system* (9) *differ by a constant, and the set of all solutions of system* (9) *is either empty or a one-dimensional linear manifold in the space of all smooth functions on the manifold M.*

In the general case, we can say that solutions of system (8) satisfy the following conditions:

(a) *the necessary condition for the existence of a solution of system* (8) *is $d\Omega = 0$,*

(b) *if ω is a solution of system* (8), *any form $\omega + d\omega'$ is also a solution of system* (8).

6.1.2. COHOMOLOGY GROUPS OF A SMOOTH MANIFOLD (THE DE RHAM COHOMOLOGY GROUPS)

Systems of the type (8) are very important for studying the structure of a manifold. It is convenient to formulate the properties of such systems within the framework of the de Rham cohomology theory.

Definition 2. An exterior differential form ω on a smooth manifold M is called *closed* if $d\omega = 0$. The form ω is called *exact* if it can be represented as $\omega = d\omega'$. The quotient space of the space of closed forms of degree k with respect to the subspace of exact forms is said to be the (*de Rham*) *cohomology group of dimension k* of the manifold M and is denoted by $H^k(M)$.

Any exact form ω is closed because $d\omega = d(d\omega') = dd(\omega) = 0$. The space of all exact forms is, therefore, a subspace in the space of closed forms. The cohomology group $H^k(M)$ is a vector space (generally, of an infinite dimension). In terms of cohomology groups, the problem of solving equation (8) is formulated as follows. Let ω be a closed exterior differential form of degree k and let $[\omega]$ denote the element of the cohomology group $H^k(M)$ equal to the coest of the form ω with respect to the subspace of exact forms. Then the following statement holds true.

Theorem 2. *Consider the equation*

$$d\omega = \Omega, \quad \deg \Omega = k + 1. \tag{14}$$

(a) *Equation* (14) *has a solution if and only if the form Ω it closed and the cohomology class $[\Omega] \in H^{k+1}(M)$ is zero.*

(b) *Any two solutions ω and ω' of* (14) *differ by a closed form, i.e. $d(\omega - \omega') = 0$. The set of all solutions of this equation is the coset of the form ω with respect to the subspace of all closed forms of degree k.*

(c) *The space of all closed forms of degree k is isomorphic to the direct sum of the space of exact forms of degree k and the cohomology group $H^k(M)$.*

Proof. If ω is a solution of equation (14), Ω is an exact form and, by definition, $[\Omega] = 0$. Conversely, if $[\Omega] = 0$, then Ω is an exact form, i.e. $\Omega = d\omega$ for a certain form ω which is just a solution of (14). Let ω and ω' be two solutions of (14), i.e. $d\omega = \Omega$ and $d\omega' = \Omega$. Then $d(\omega - \omega') = d\omega - d\omega' = \Omega - \Omega = 0$, i.e. $\omega - \omega'$ is a closed form. Hence, any solution ω' of equation (14) can be obtained by adding a closed form to ω. Let $\Omega_k(M)$ denote the linear space of all exterior differential forms of degree k on a manifold M. Then the gradient d is the linear mapping

$$d\colon \Omega_k(M) \to \Omega_{k+1}(M) \tag{15}$$

from the space of forms of degree k into the space of forms of degree $k + 1$. The space of closed forms of degree k coincides with the kernel $\ker d \subset \Omega_k(M)$ of mapping (15). Similarly, the gradient d maps the space $\Omega_{k-1}(M)$ into $\Omega_k(M)$:

$$d\colon \Omega_{k-1}(M) \to \Omega_k(M). \tag{16}$$

Then the space of exact forms of degree k coincides with the image of mapping (16): $\operatorname{Im} d = d(\Omega_{k-1}(M)) \subset \ker d \subset \Omega_k(M)$. Hence, the k-dimensional cohomology group $H^k(M)$ coincides with the quotient space $\ker d/\operatorname{Im} d$. In the notation of the latter space the same symbol d stands for two distinct mappings (15) and (16). This does not, however, lead to confusion, because the subspaces we are dealing with are considered in one space of forms of degree k, $\Omega_k(M)$. Let us consider in the linear space $\ker d$ the cofactor H' to the subspace $\operatorname{Im} d \subset$

ker d. Then the space ker d is decomposed into the direct sum of its subspaces: $\ker d = \operatorname{Im} d \oplus H'$.

We now demonstrate that the space H' is isomorphic to the cohomology group $H^k(M)$. Indeed, let $\varphi\colon H' \to H^k(M)$ be a mapping which associates with a closed form $\omega \in H'$ its coset $[\omega] \in H^k(M)$. If $\varphi[\omega] = 0$, then $[\omega] = 0$, i.e. ω is an exact form. This means that $\omega \in \operatorname{Im} d$, and since the subspaces Im d and H' intersect with respect to zero element, we have $\omega = 0$, so that the mapping φ is a monomorphism. Let $x \in H^k(M)$ be an arbitrary element, then x is the coset of some form $\omega \in \ker d$ with respect to the subspace Im d, $x = [\omega]$. Since the space ker d can be decomposed into the direct sum of the subspaces Im d and H', the form ω can also be decomposed into the sum $\omega = d\Omega + \omega'$, $d\Omega \in \operatorname{Im} d$, $\omega' \in H'$, so that $x = [\omega] = [\omega'] = \varphi(\omega')$. Thus, φ is an epimorphism. We have proved that φ is an isomorphism, which completes the proof of Theorem 2.

Examples

1. Let a manifold M be represented by an open real interval, $M = (a, b)$. Calculate the cohomology groups of this interval. Since the manifold M is one-dimensional, only the spaces of the forms of degrees 0 and 1 are different from zero. Consider first the space $\Omega_0(M)$ of the forms of degree 0. Any such form is a smooth function on the interval (a, b), and its gradient df is written as $df = \frac{df}{dx}(x)\,dx$, where x is the Cartesian coordinate on (a, b). Therefore, the space of closed forms of degree 0, i.e. ker d, consists of all functions such that $\frac{df}{dx}(x) \equiv 0$, i.e. ker d consists of constant functions. Hence, the space ker d is isomorphic to the one-dimensional space $\mathbf{R}^1$. The space $\Omega_0(M)$ does not contain exact forms and, therefore, $H^0(M) = \ker d = \mathbf{R}^1$. Let us now consider the space $\Omega_1(M)$ of the forms of degree 1. Since any form of degree 2 on a one-dimensional manifold vanishes, ker d coincides with $\Omega_1(M)$. Calculate Im d. Let $\omega \in \Omega_1(M)$ be an arbitrary differential form of degree 1. In a local coordinate system it is written as $\omega = g(x)\,dx$. If $df = \omega$, then $\frac{\partial f}{\partial x}(x)\,dx \equiv g(x)\,dx$, i.e. $\frac{\partial f}{\partial x} = g$. Hence, the function f can be defined by

$$f(x) = \int_c^x g(x)\,dx, \quad x \in (a, b), \quad a < c < b. \tag{17}$$

Thus, any form $\omega \in \Omega_1(M)$ can be represented as $\omega = df$ (for an appropriate function f). This means that $\operatorname{Im} d = \Omega_1(M) = \ker d$. Then the one-dimensional cohomology group $H^1(M)$ defined as the quotient space $\ker d/\operatorname{Im} d$ is zero, $H^1(M) = 0$. For other dimensions

$k \geqslant 2$ the cohomology groups $H^k(M)$ vanish because even for $k \geqslant 2$ the spaces of the forms of degree k are zero on a one-dimensional manifold M. Thus, for $k \geqslant 1$ we have $H^0(M) = \mathbf{R}^1$, $H^k(M) = 0$.

2. Let $M = S^1$ be a one-dimensional circle. As in Example 1, we find that the zero-dimensional cohomology group $H^0(S^1) = \mathbf{R}^1$ and the groups $H^k(s^1) = 0$ for $k \geqslant 2$. It remains to calculate the one-dimensional cohomology group $H^1(S^1)$. Since S^1 is one-dimensional, the kernel $\ker d$ coincides with $\Omega_1(S^1)$. Let φ be the local angular parameter of the circle S^1. Solve the equation $df = \omega$ for a certain form $\omega = g(\varphi)\,d\varphi$. The function $f(\varphi)$ must be 2π-periodic in the parameter φ. Using formula (17), we obtain $f(\varphi) = \int_0^{\varphi} g(\varphi)\,d\varphi$ and $\int_0^{2\pi} g(\varphi)\,d\varphi = 0$. Thus, not every form $\omega \in \Omega_1(S^1)$ is contained in the gradient image $\operatorname{Im} d$, but only the form satisfying the condition $\int_0^{2\pi} g(\varphi)\,d\varphi = 0$. We now demonstrate that the space of the forms $\Omega_1(S^1)$ can be decomposed into the direct sum of its subspaces: $\Omega_1(S^1) = \operatorname{Im} d \oplus \mathbf{R}^1$, the second term consisting of forms with constant coefficient, $g(\varphi) \equiv \text{const}$. Indeed, let $\omega = g(\varphi)\,d\varphi$ be an arbitrary form, $c = \frac{1}{2\pi}\int_0^{2\pi} g(\varphi)\,d\varphi$. Then the form $\omega' = g'(\varphi)\,d\varphi = (g(\varphi) - c)\,d\varphi$ lies in $\operatorname{Im} d$ because $\int_0^{2\pi} g'(\varphi)\,d\varphi = 0$. On the other hand, any form $\omega = g(\varphi)\,d\varphi$ with constant coefficient $g(\varphi) \equiv c = \text{const}$ does not lie in $\operatorname{Im} d$ for $c \neq 0$, since $\int_0^{2\pi} c\,d\varphi = 2\pi c \neq 0$. Then the factor group $H^1(S^1) = \ker d/\operatorname{Im} d = \Omega_1(S^1)/\operatorname{Im} d = \mathbf{R}^1$. Thus, we finally obtain $H^0(S^1) = H^1(S^1) = \mathbf{R}^1$, $H^k(S^1) = 0$ for $k \geqslant 2$.

6.1.3. HOMOTOPIC PROPERTIES OF COHOMOLOGY GROUPS

Cohomology groups of smooth manifolds possess some properties which may prove useful in describing these groups. We note first of all that each smooth mapping of a manifold induces the inverse mapping of an exterior differential form, namely, $f\colon M_1 \to M_2$, $f^*\colon \Omega_k(M_2) \to \Omega_k(M_1)$.

Definition 3. Let $f\colon M_1 \to M_2$ be a smooth mapping and let $\omega \in \Omega_k(M_2)$ be an exterior differential form of degree k on the manifold M_2. The *inverse image* $f^*(\omega)$ *of the form* ω is an exterior differential form on M_1 defined by $f^*(\omega)(\xi_1, \ldots, \xi_k) = \omega(df(\xi_1), \ldots, df(\xi_k))$, where $\xi_1, \ldots, \xi_k \in T_P(M_1)$ are tangent vectors at a point P of M_1 and $df(\xi_1), \ldots, df(\xi_k) \in T_{f(P)}(M_2)$ are their images under the mapping df.

The mapping $f^*\colon \Omega_k(M_2) \to \Omega_k(M_1)$ is, apparently, a linear mapping of a linear space.

Theorem 3. *The inverse image of a differential form satisfies the following conditions*:

$$\text{(a) } (gf)^* = f^*g^*,\ f\colon M_1 \to M_2,\ g\colon M_2 \to M_3, \tag{18}$$

$$\text{(b) } f^*d = df^*, \tag{19}$$

(c) *the mapping* f^* *sends* $\ker d$ *into* $\ker d$ *and also cohomology classes into cohomology classes*: $f^*\colon H^k(M_2) \to H^k(M_1)$,

(d) *if in local coordinates the mapping* f *is written as a system of functions* $y^i = f^i(x^1, \ldots, x^n)$, *and the form*

$$\omega = \Sigma\omega_{i_1 \ldots i_k}(y^1, \ldots, y^m)\, dy^{i_1} \wedge \ldots \wedge dy^{i_k},$$

then

$$\begin{aligned} f^*(\omega) = \sum \omega_{i_1 \ldots i_k}&(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)) \\ &\cdot df^{i_1}(x^1, \ldots, x^n) \wedge \ldots \wedge df^{i_k}(x^1, \ldots, x^n). \end{aligned} \tag{20}$$

Relation (20) is a generalization of the so-called invariance property of the differential of a function, which states that the differential of a function $y = f(x)$ does not depend on whether x is an independent variable or, in turn, is a function of another variable: $dy = \frac{df}{dx}(x)\, dx$.

We begin the proof of Theorem 3 by verifying relation (20). Let $\xi_1 = (\xi_1^i), \ldots, \xi_k = (\xi_k^i)$ be tangent vectors at a point P of the manifold M_1 referred to a local coordinate system $(x^1, \ldots, x^n)$. By definition, we have

$$f^*(\omega)(\xi_1, \ldots, \xi_k) = \omega(df(\xi_1), \ldots, df(\xi_k)). \tag{21}$$

The components of the vectors $\eta_i = df(\xi_i)$ in a local coordinate system $(y^1, \ldots, y^m)$ of the manifold M_2 are of the form

$$\eta_i^j = \sum_k \frac{\partial f^j}{\partial x^k}\xi_i^k = df^j(\xi_i).$$

Thus,

$$
\begin{aligned}
f^*(\omega)(\xi_1, \ldots, \xi_k) &= \sum \omega_{j_1, \ldots, j_k} \eta_1^{j_1} \eta_2^{j_2} \ldots \eta_k^{j_k} \\
&= \sum \omega_{j_1 \ldots j_k} df^{j_1}(\xi_1) \ldots df^{j_k}(\xi_k) \\
&= \left(\sum \omega_{j_1 \ldots j_k} df^{j_1} \wedge \ldots \wedge df^{j_k}\right)(\xi_1, \ldots, \xi_k).
\end{aligned}
$$

Formula (20) is proved. Relation (18) directly follows from (20) in any local coordinate system. Let us prove relation (19). According to (20), if

$$\omega = \sum \omega_{i_1 \ldots i_k}(y^1, \ldots, y^m)\, dy^{i_1} \wedge \ldots \wedge dy^{i_k},$$

then

$$f^*(\omega) = \sum \omega_{i_1 \ldots i_k}(f^1(x^1, \ldots, x^n), \ldots)\, df^{i_1} \wedge \ldots \wedge df^{i_k}.$$

We have therefore

$$
\begin{aligned}
d\omega &= \sum d\omega_{i_1 \ldots i_k}(y^1, \ldots, y^m) \wedge dy^{i_1} \wedge \ldots \wedge dy^{i_k} \\
&= \sum \frac{\partial \omega_{i_1 \ldots i_k}}{\partial y^j}(y^1, \ldots, y^m)\, dy^j \wedge dy^{i_1} \wedge \ldots \wedge dy^{i_k},
\end{aligned}
$$

$$
\begin{aligned}
f^* d\omega &= \sum \frac{\partial \omega_{i_1 \ldots i_k}}{\partial y^j}(f^1(x^1, \ldots, x^n), \ldots)\, df^j \wedge df^{i_1} \wedge \ldots \wedge df^{i_k} \\
&= \sum d(\omega_{i_1 \ldots i_k})(f^1(x^1, \ldots, x^n), \ldots)) \wedge df^{i_1} \wedge \ldots \wedge df^{i_k}. \quad (22)
\end{aligned}
$$

On the other hand,

$$df^*(\omega) = \sum d(\omega_{i_1 \ldots i_k}(f^1(x^1, \ldots, x^n) \ldots)) \wedge df^{i_1} \wedge \ldots \wedge df^{i_k}, \quad (23)$$

i.e. the right-hand sides of (22) and (23) coincide. Finally, condition (c) follows from (b). Indeed, if $\omega \in \ker d \subset \Omega_k(M_2)$, then $d\omega = 0$ and $df^*(\omega) = f^*d(\omega) = 0$, i.e. $f^*(\omega) \in \ker d \subset \Omega_k(M_1)$. The cohomology class $[\omega] \in H^k(M_2)$ is the coset of the form ω with respect to the subspace $\operatorname{Im} d \subset \ker d \subset \Omega_k(M_2)$. The image of this coset under the mapping f^* consists of the forms $f^*(\omega + d\omega')$, $\omega' \in \Omega_{k-1}(M_2)$. According to (19), all such forms are expressed as $f^*(\omega + d\omega') = f^*(\omega) + df^*(\omega')$, i.e. they lie in the coset of the form $f^*(\omega)$ with respect to the subspace $\operatorname{Im} d \subset \ker d \subset \Omega_k(M_1)$. Thus, the homomorphism $f^*: H^k(M_2) \to H^k(M_1)$ of the cohomology groups of M_1 and M_2 is defined correctly. By virtue of condition (a), the relation $(gf)^* = f^*g^*$ also holds true for mappings of cohomology groups. The theorem is proved.

Theorem 3 is useful in many respects. For example, if two manifolds M_1 and M_2 are diffeomorphic, their cohomology groups are

isomorphic. Indeed, if $f: M_1 \to M_1$ is an identity diffeomorphism, the homomorphism $f^*: H^k(M_1) \to H^k(M_1)$ is an identity isomorphism. Thus, by choosing an arbitrary diffeomorphism $\varphi: M_1 \to M_2$ and calculating the inverse diffeomorphism $\psi: M_2 \to M_1$, we obtain two possible mappings $\psi\varphi: M_1 \to M_1$ and $\varphi\psi: M_2 \to M_2$, both being identity diffeomorphisms: $\psi\varphi = 1_{M_1}$ and $\varphi\psi = 1_{M_2}$. Using then Theorem 3, we find that $\varphi^*\psi^* = 1_{H^k(M_1)}$, $\psi^*\varphi^* = 1_{H^k(M_2)}$, i.e. the homomorphisms φ^* and ψ^* are mutually inverse and, therefore, φ^* (as well as ψ^*) is an isomorphism of cohomology groups.

It turns out, however, that the cohomology homomorphisms f^*: $H^k(M_2) \to H^k(M_1)$ induced by a smooth mapping $f: M_1 \to M_2$ possess even more stringent properties.

Definition 4. Let X, Y be arbitrary topological spaces and let I be a unit segment of real numbers. The continuous mapping f: $X \times I \searrow Y$ is called a *homotopy between the mappings* $f_0 = f \mid (X \times \{0\})$ and $f_1 = f \mid (X \times \{1\})$. If X and Y are smooth manifolds and f is a smooth mapping, the homotopy f is called *smooth*, and the mappings f_0 and f_1 are called *homotopic mappings*.

Remark. If X is a smooth manifold, the Cartesian product $X \times I$ is not a manifold (rigorous definitions are given in the next section). Yet, the concept of a smooth mapping holds true for a Cartesian product. Indeed, if $U_\alpha \subset X$ is a chart and $(x_\alpha^1, \ldots, x_\alpha^n)$ is a local coordinate system, the mapping f on the set $U_\alpha \times I$ is represented as a function of $(n+1)$ variables, $f = f(x_\alpha^1, \ldots, x_\alpha^n, t)$, $t \in I = [0, 1]$. The function $f(x_\alpha^1, \ldots, x_\alpha^n, t)$ is said to be *smooth* if in the neighbourhood of each point $(x_\alpha^1, \ldots, x_\alpha^n, t)$, $0 < t < 1$, it is smooth of class C^∞, and all its partial derivatives can be extended continuously to the boundary points $(x_\alpha^1, \ldots, x_\alpha^n, 0)$ or $(x_\alpha^1, \ldots, x_\alpha^n, 1)$. This definition does not depend on the choice of the local coordinate system, and the restrictions $f \mid (X \times \{0\})$, $f \mid X \times \{1\})$ are smooth mappings of X into Y.

Theorem 4. *Homotopic smooth mappings f_0, f_1: $M_1 \to M_2$ induce the same homomorphism of cohomology groups $f_0^* = f_1^*$: $H^k(M_2) \to H^k(M_1)$.*

Proof. Let us construct the linear mapping of the spaces of exterior differential forms D: $\Omega_k(M_2) \to \Omega_{k-1}(M_1)$ such that the identity

$$(f_0^* - f_1^*)(\omega) = (dD \pm Dd)(\omega) \tag{24}$$

is satisfied for any form $\omega \in \Omega_k(M_2)$. If ω is a closed form representing the cohomology class $[\omega] \in H^k(M_2)$, the forms $f_0^*(\omega)$ and $f_1^*(\omega)$ are also closed and represent the cohomology classes $f_0^*([\omega])$ and $f_1^*([\omega])$. By virtue of (24), the difference $f_0^*(\omega) - f_1^*(\omega)$ is equal to $(dD \pm Dd)\,\omega = d(D\omega) \pm D(d\omega) = d(D\omega)$ because $d\omega = 0$. The forms $f_0^*(\omega)$ and $f_1^*(\omega)$ belong, therefore, to the same coset

with respect to the subspace of exact forms, i.e. they represent the same cohomology class $[f_0^*(\omega)] = [f_1^*(\omega)]$. Thus, $f_0^*([\omega]) = f_1^*([\omega])$.

It remains to construct the mapping D satisfying identity (24). Since the mappings f_0 and f_1 are homotopic, there exists a smooth mapping $F\colon M_1 \times I \to M_2$ such that $F(P, 0) = f_0(P)$ and $F(P, 1) = f_1(P)$. Let $\omega \in \Omega_k(M_2)$ be an arbitrary form of degree k and $\Omega = F^*(\omega)$. The form Ω is defined at least on the manifold $M_1 \times (0, 1)$, and in any local coordinate system all the coefficients of the form Ω can be extended continuously to $M_1 \times [0, 1]$. A form Ω of degree k on the manifold $M_1 \times I$ is said to be *independent of dt* if this form vanishes on any system of vectors $\left(\xi_1, \ldots, \xi_{k-1}, \frac{\partial}{\partial t}\right)$. In the local coordinate system $(x^1, \ldots, x^n, t)$ the condition that the form Ω does not depend on dt means that it can be decomposed into the sum $\Omega = \Sigma\Omega_{i_1\ldots i_k}(x^1, \ldots, x^n, t)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, i.e. the decomposition does not include terms with dt. In this case the form Ω can be interpreted as a family of exterior differential forms $\Omega_{M_1}(t)$ of degree k (on the manifold M_1) continuously dependent on the parameter t. If $\varphi_t\colon M_1 \to M_1 \times I$ is an embedding $\varphi_t(P) = (P, t)$, then $\Omega_{M_1}(t) = \varphi_t^*(\Omega)$.

Lemma 1. *Any exterior differential form Ω of degree k on the manifold $M_1 \times I$ admits the unique representation*

$$\Omega = \Omega_1 + \Omega_2 \wedge dt, \tag{25}$$

where the forms Ω_1 and Ω_2 do not depend on dt.

It is sufficient to *prove* the lemma in each chart U_α of the manifold M_1. Indeed, the independence of a form of dt is invariant relative to the choice of the chart and is valid at each point. Therefore, decompositions of type (25) realized in each chart coincide in the intersections of the charts. Since the condition of independence of dt is invariant under linear operations over forms, it suffices to verify the lemma for special forms $\Omega = f(x^1, \ldots, x^n, t)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ or $\Omega = f(x^1, \ldots, x^n, t)\, dx^{i_1} \wedge \ldots \wedge dx^{i_{k-1}} \wedge dt$ for which the lemma is obvious.

Let us now define the mapping D. We put

$$D(\omega) = \int_0^1 \Omega_{2, M_1}(t)\, dt, \tag{26}$$

where $\Omega = F^*(\omega)$ and $\Omega = \Omega_1 + \Omega_2 \wedge dt$ is the decomposition by formula (25). If the form Ω on $M_1 \times I$ does not depend on dt, its gradient $d\Omega$ can also be decomposed by (25) into the sum $d\Omega = d_x\Omega_{M_1} + \frac{\partial}{\partial t}\Omega_{M_1} \wedge dt$. Here d_x denotes the gradient of Ω_{M_1} on the manifold M_1. We now prove formula (24). The left-hand side of (24)

can be obtained as follows. First, we calculate the form $\Omega = F^*(\omega)$ and then calculate its restrictions to the submanifolds $M_1 \times \{0\}$ and $M_2 \times \{1\}$ under the embeddings φ_0: $M_1 \to M_1 \times I$ and φ_1: $M_1 \to M_1 \times I$. In a local coordinate system the embeddings φ_0 and φ_1 are defined by the functions

$$\varphi_0: \begin{cases} x^1 = x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^n = x^n \\ t \equiv 0 \end{cases}, \quad \varphi_1: \begin{cases} x^1 = x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^n = x^n \\ t \equiv 1 \end{cases}.$$

In other words, $t = 0$ or 1 and $dt = 0$ should be substituted into the form Ω. This means that

$$f_0^*(\omega) = \Omega_{1, M_1}(0), \quad f_1^*(\omega) = \Omega_{1, M_1}(1). \tag{27}$$

To find the right-hand side of (24), we calculate successively $Dd(\omega)$ and $dD(\omega)$. We have

$$\begin{aligned} F^*(d\omega) &= dF^*(\omega) = d(\Omega_1 + \Omega_2 \wedge dt) \\ &= d_x\Omega_{1, M_1}(t) + \frac{\partial}{\partial t}\Omega_{1, M_1}(t) \wedge dt + d_x\Omega_{2, M_1} \wedge dt, \end{aligned}$$

whence

$$\begin{aligned} Dd(\omega) &= \int_0^1 \left(\frac{\partial}{\partial t}\Omega_{1, M_1}(t) + d_x\Omega_{2, M_1}(t)\right) dt \\ &= \Omega_{1, M_1}(1) = \Omega_{1, M_1}(0) + d_x \int_0^1 \Omega_{2, M_1}(t)\, dt. \end{aligned} \tag{28}$$

On the other hand

$$dD(\omega) = d_x \int_0^1 \Omega_{2, M_1}(t)\, dt. \tag{29}$$

Comparison of (27) and (28), (29) yields $f_1^*(\omega) - f_0^*(\omega) = (Dd - dD)(\omega)$, which is what was required.

Corollary. *If $M = \mathbf{R}^n$ is an n-dimensional Euclidean space, then $H^0(M) = \mathbf{R}^1$ and $H^k(\mathbf{R}^n) = 0$, $k \geqslant 1$.*

Proof. For $n = 0$ the manifold M consists of a single point and the statement is obvious, since a single-point manifold does not have non-trivial forms of degree higher than zero. And the forms of

degree 0 are functions on M, i.e. just real numbers. Thus, $\Omega_0(M) = \mathbf{R}^1$, the kernel of the gradient, coincides with $\Omega_0(M)$, and the image of the gradient vanishes. Hence, $H^0(M) = \ker d/\operatorname{Im} d = \Omega_0(M)/0 = \Omega_0(M) = \mathbf{R}^1$. Let now $n > 0$ and let $M_0 = \mathbf{R}^0$ be a single-point manifold. Consider two mappings φ: $M_0 \to M$, $\varphi(M_0) = 0 \in M$, and ψ: $M \to M_0$, $\psi(M) = 0 = M_0$. Thus, the mapping φ sends the only point of the manifold M_0 into zero vector, and the mapping ψ sends the entire Euclidean space $M = \mathbf{R}^n$ into one point M_0. Let us consider two possible mappings $\psi\varphi$: $M_0 \to M_0$ and $\varphi\psi$: $M \to M$. The mapping $\psi\varphi$ is, apparently, an identity transformation of the single-point manifold M_0, while $\varphi\psi$ maps the entire Euclidean space M into zero vector. We now demonstrate that the mapping $\varphi\psi$ is homotopic to an identity mapping of Euclidean space. Construct the homotopy in the explicit form, F: $M \times I \to M$, $F(\mathbf{x}, t) = t\mathbf{x}$, $\mathbf{x} \in M = \mathbf{R}^n$, $t \in [0, 1] = I$. For $t = 1$ we obtain the identity mapping $f_1(\mathbf{x}) = \mathbf{x} = F(\mathbf{x}, 1)$ and for $t = 0$ we have $\varphi\psi(\mathbf{x}) = \mathbf{0} = F(\mathbf{x}, 0)$. According to Theorem 4, the mapping $\varphi\psi$ and the identity mapping both induce the same homomorphism of cohomology groups and, therefore, the homomorphisms $(\psi\varphi)^*$: $H^k(M_0) \to H^k(M_0)$ and $(\varphi\psi)^*$: $H^k(M) \to H^k(M)$ are identical isomorphisms of groups. Since $(\varphi\psi)^* = \psi^*\varphi^*$ and $(\psi\varphi)^* = \varphi^*\psi^*$, the homomorphisms φ^* and ψ^* are mutually inverse isomorphisms of groups. Hence, a Euclidean space has the same cohomology groups $H^k(\mathbf{R}^n)$ as a single-point space, i.e. $H^0(\mathbf{R}^n) = \mathbf{R}^1$, $H^k(\mathbf{R}^n) = 0$, $k \geqslant 1$. The corollary is proved.

This corollary has another formulation known as Poincaré's lemma.

Theorem 5 (Poincaré's lemma). *Any closed form Ω on a manifold M is exact in a sufficiently small neighbourhood of each point $P \in M$, i.e. $\Omega = d\omega$.*

Proof. Consider a neighbourhood U of a point $P \in M$ homeomorphic to Euclidean space $\mathbf{R}^n$. Then we have $H^k(U) = 0$ for $k \geqslant 1$. Hence, if Ω is a closed form on M, the cohomology class $[\Omega]$ vanishes, which means that the form Ω is exact in the neighbourhood U. Theorem 5 is proved.

6.2. INTEGRATION OF EXTERIOR FORMS

Integration of an exterior differential form is an analogue of the operation inverse to differentiation of a function. Let us consider an example. Let an exterior differential form ω of degree n be given in a bounded domain V of a Euclidean space $\mathbf{R}^n$. Then in the local coordinate system $(x^1, \ldots, x^n)$ the form ω is represented by

$$\omega = f(x^1, \ldots, x^n)\, dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n. \qquad (1)$$

In some other coordinate system $(y^1, \ldots, y^n)$ the form will be written, according to the transformation law of tensor components, as

$$\omega = f(x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)) \det\left(\frac{\partial x^i}{\partial y^j}\right) dy^1 \wedge \ldots \wedge dy^n. \tag{2}$$

We see that this formula resembles the formula for the transformation of an integrand function under coordinate substitution (in the integral calculus of functions of many variables). Namely, the values of the integral of a function f over a volume V in different coordinate systems are related by

$$\int_V f(x^1, \ldots, x^n)\, dx^1 \ldots dx^n$$
$$= \int_V f(x^1(y^1, \ldots, y^n), \ldots) \left|\det\left(\frac{\partial x^i}{\partial y^j}\right)\right| dy^1 \ldots dy^n. \tag{3}$$

Formula (3) shows that it is not in fact the function f that is integrated over the volume V, but a certain geometric object whose components are transformed under coordinate substitution $(x^1, \ldots, x^n) \to (y^1, \ldots, y^n)$ via multiplication by the modulus of the Jacobian of this substitution. This argument can be used to define the integral of an exterior differential form.

6.2.1. THE INTEGRAL OF A DIFFERENTIAL FORM OVER A MANIFOLD

Let us now return to formula (3) and extend it to an arbitrary smooth manifold M with boundary. A set $X \subset M$ is said to be of *zero measure*, $\mu(X) = 0$, if for any chart U_α and the coordinate homeomorphism φ_α: $U_\alpha \subset V_\alpha \subset \mathbf{R}^n_+$ the set $\varphi_\alpha(X \cap U_\alpha) \subset \mathbf{R}^n_+$ is of zero measure in Euclidean space $\mathbf{R}^n$. The definition of a set of zero measure is correct and does not depend on the choice of the atlas. This can be proved by showing that if f: $V_1 \to V_2$ is a smooth homeomorphism of the domain $V_1 \subset \mathbf{R}^n$ onto the domain $V_2 \subset \mathbf{R}^n$, then the image $f(X)$ of any set $X \subset V_1$ of zero measure is also a 0-measure set. Indeed, if $X \subset V_1$ is a set of zero measure, $\mu(X) = 0$, it can be represented as the union of not more than a countable number of compact subsets: $X = \bigcup_{i=1}^{\infty} X_i$, $\mu(X_i) = 0$. Then, $\mu(f(X)) \leqslant \sum_{i=1}^{\infty} \mu(f(X_i))$, and it is therefore sufficient to prove that $\mu(f(X_i)) = 0$.

In a local coordinate system $(x^1, \ldots, x^n)$ the mapping f is represented as the set of smooth functions $y^i = f^i(x^1, \ldots, x^n)$. Thus,

$$\mu(f(X_i)) \leqslant \int\limits_{X_i} \left|\det\left(\frac{\partial f^k}{\partial x^j}\right)\right| dx^1 \ldots dx^n \leqslant C \int\limits_{X_i} dx^1 \ldots dx^n = C\mu(X_i),$$

where C is a constant satisfying

$$\det\left|\left(\frac{\partial f^k}{\partial x^j}\right)\right| \leqslant C$$

on the compact X_i. A set of zero measure on a manifold obeys the ordinary properties of a set in a Euclidean space:

(a) $\mu\left(\bigcup\limits_{i=1}^{\infty} X_i\right) \leqslant \sum\limits_{i=1}^{\infty} \mu(X_i) = 0,$

(b) the image $f(Y)$ of a smooth $(n-1)$-dimensional manifold Y obtained by the mapping $f: Y \to M$ is of zero measure.

Let us now define the integral of an exterior differential form ω of degree n on an n-dimensional oriented manifold M when the form ω has a compact support. Let $U \subset M$ be an arbitrary open domain. If U is entirely contained in some chart U_α with coordinates $(x_\alpha^1, \ldots, x_\alpha^n)$, the form ω can be expressed in the chart U_α as $\omega = f_\alpha(x_\alpha^1, \ldots, x_\alpha^n)\, dx_\alpha^1 \wedge \ldots \wedge dx_\alpha^n$. We put then

$$\int\limits_U \omega = \int \underset{U}{\ldots} \int f_\alpha(x_\alpha^1, \ldots, x_\alpha^n)\, dx_\alpha^1 \ldots dx_\alpha^n. \tag{4}$$

The right-hand side of (4) does not depend on the choice of the local coordinate system $(x_\alpha^1, \ldots, x_\alpha^n)$. Indeed, if $U \subset U_\alpha \cap U_\beta$, it follows from formula (2)

$$\begin{aligned}\omega &= f_\alpha(x_\alpha^1(x_\beta^1, \ldots, x_\beta^n), \ldots, x_\alpha^n(x_\beta^1, \ldots, x_\beta^n)) \\ &\quad \times \det\left(\frac{\partial x_\alpha^i}{\partial x_\beta^j}\right) dx_\beta^1 \wedge \ldots \wedge dx_\beta^n \\ &= f_\beta(x_\beta^1, \ldots, x_\beta^n)\, dx_\beta^1 \wedge \ldots dx_\beta^n.\end{aligned}$$

Using (3) and taking into account that for an oriented manifold $\det\left(\frac{\partial x_\alpha^i}{\partial x_\beta^j}\right) > 0$, we obtain

$$\begin{aligned}&\int \underset{U}{\ldots} \int f_\alpha(x_\alpha^1, \ldots, x_\alpha^n)\, dx_\alpha^1 \ldots dx_\alpha^n \\ &\qquad = \int \underset{U}{\ldots} \int f_\beta(x_\beta^1, \ldots, x_\beta^n)\, dx_\beta^1 \ldots dx_\beta^n.\end{aligned}$$

Definition 1. Let M be an n-dimensional smooth oriented manifold with boundary and let ω be an exterior differential form of degree n with compact support. Furthermore, let $\{U_{\alpha_1}, \ldots, U_{\alpha_N}\}$ be a finite set of charts covering the support of the form ω. We put then

$$\int_M \omega = \sum_{i=1}^{N} \int_{(U_{\alpha_i} \setminus \bigcup_{j=1} U_{\alpha_j})} \omega_{i-1}. \tag{5}$$

Definition 1 does not depend on the choice of the charts $U_{\alpha_1}, \ldots, U_{\alpha_N}$. To demonstrate this, we give another definition of the integral of a form ω.

Definition 1′. Let M be an n-dimensional smooth oriented manifold with boundary, let ω be an exterior differential form of degree n with compact support, and let $\{\varphi_\alpha\}$ be a partition of unity subordinate to the atlas $\{U_\alpha\}$. Then

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \varphi_\alpha \omega. \tag{6}$$

Definition 1′ is more convenient, for it relies upon integration only over an open domain in a Euclidean space, whereas in formula (5) integration is performed over a more complicated domain. On the other hand, in Definition 1′ it is required to prove that the right-hand side of (6) does not depend on the choice of the partition of unity.

Without loss of generality, we may assume that the atlas U_α is finite, $1 \leqslant \alpha \leqslant N$. Let $\{\varphi'_\alpha\}$ be another partition of unity. Putting $\psi_\alpha = \varphi_\alpha - \varphi'_\alpha$, we obtain $\sum_{\alpha=1}^{N} \psi_\alpha \equiv 0$. It is necessary to demonstrate that

$$\sum_{\alpha=1}^{N} \int_{U \cap U_\alpha} \psi_\alpha \omega = 0. \tag{7}$$

We have $\psi_N = -\sum_{\alpha=1}^{N-1} \psi_\alpha$; since $\operatorname{supp} \psi_N \subset U_N$, there exists another function χ, which is identically 1 on $\operatorname{supp} \psi_N$, so that $\operatorname{supp} \chi \subset U_N$. Then, $\chi(P)\,\psi_N(P) \equiv \psi_N(P)$ and therefore $\psi_N = -\sum_{\alpha=1}^{N-1} \chi\psi_\alpha$, $\operatorname{supp} \chi\psi_\alpha \subset U_N \cap U_\alpha$. It follows that

$$\sum_{\alpha=1}^{N} \int_{U_\alpha} \psi_\alpha \omega = \int_{U_N} \psi_N \omega + \sum_{\alpha=1}^{N-1} \int_{U_\alpha} \psi_\alpha \omega = -\sum_{\alpha=1}^{N-1} \int_{U_N} \chi\psi_\alpha \omega + \sum_{\alpha=1}^{N-1} \int_{U_\alpha} \psi_\alpha \omega.$$

Since $\operatorname{supp} \chi\psi_\alpha\omega \subset U_N \cap U_\alpha$, we have $\int\limits_{U_N} \chi\psi_\alpha\omega = \int\limits_{U_\alpha} \chi\psi_\alpha\omega$, whence

$$\sum_{\alpha=1}^{N} \int\limits_{U_\alpha} \psi_\alpha\omega = \sum_{\alpha=1}^{N-1} \int\limits_{U_\alpha} (\psi_\alpha - \chi\psi_\alpha)\,\omega. \tag{8}$$

The right-hand side of (8) includes $(N-1)$ terms and coincides with (7) in which the functions $(\psi_\alpha - \chi\psi_\alpha)$ are substituted for ψ_α. We have

$$\sum_{\alpha=1}^{N-1} (\psi_\alpha - \chi\psi_\alpha) = \sum_{\alpha=1}^{N-1} \psi_\alpha - \chi \sum_{\alpha=1}^{N-1} \psi_\alpha = \sum_{\alpha=1}^{N-1} \psi_\alpha + \chi\psi_N$$
$$= \sum_{\alpha=1}^{N-1} \psi_\alpha + \psi_N = \sum_{\alpha=1}^{N} \psi_\alpha = 0.$$

Thus, formula (7) is proved by induction with respect to the number of terms N. Let now there be given two atlases $\{U_\alpha\}$ and $\{U'_\beta\}$ with the corresponding subordinate partitions of unity $\{\varphi_\alpha\}$ and $\{\varphi'_\beta\}$. Each partition can be completed with zero functions to the partition of unity of a new atlas equal to the union $\{U_\alpha\} \cup \{U^\beta\}$. The zero functions do not alter the right-hand side of (6) and formula (7) implies that the right-hand sides of (6) coincide for the partitions of unity $\{\varphi_\alpha\}$ and $\{\varphi'_\beta\}$.

Proposition 1. (a) *Definitions of the integral by formulas* (5) *and* (6) *coincide*,

(b) $\int\limits_M (\lambda_1\omega_1 + \lambda_2\omega_2) = \lambda_1 \int\limits_M \omega_1 + \lambda_2 \int\limits_M \omega_2$, $\lambda_1, \lambda_2 \in \mathbf{R}^1$,

(c) *if the orientation of M' is opposite to the orientation of M, then* $\int\limits_{M'} \omega = -\int\limits_M \omega$.

Proof. Property (b) is obvious both for formula (5) and formula (6). It is sufficient, therefore, to prove that (5) and (6) coincide for forms with a sufficiently small support. If $\operatorname{supp} \omega \subset U_{\alpha i_0}$, it follows from (4) that $\int\limits_U \omega = \int\limits_{U \cap U_{\alpha_{i_0}}} \omega$ and in formula (5) we have

$$\sum_{i=1}^{N} \int\limits_{(U_{\alpha_i} \setminus \bigcup_{j=1}^{i-1} U_{\alpha_j})} \omega = \sum_{i=1}^{N} \int\limits_{U_{\alpha_{i_0}} \cap (U_{\alpha_i} \setminus \bigcup_{j=1}^{i-1} U_{\alpha_j})} \omega.$$

Since all the sets $U_{\alpha_{i_0}} \cap \left(U_{\alpha_i} \setminus \bigcup_{j=1}^{i-1} U_{\alpha_j}\right)$ belong to the same chart and do not intersect pairwise, the sum of the integrals is equal to the

integral over the union

$$\bigcup_{i=1}^{N}\left(U_{\alpha_{i_0}}\cap\left(U_{\alpha_i}\setminus\bigcup_{j=1}^{i-1}U_{\alpha_j}\right)\right)=U_{\alpha_{i_0}}\cap\left(\bigcup_{i=1}^{N}\left(U_{\alpha_i}\setminus\bigcup_{j=1}^{i-1}U_{\alpha_j}\right)\right)$$
$$=U_{\alpha_{i_0}}\cap M=U_{\alpha_{i_0}}.$$

Hence, the right-hand side of (5) is of the form $\int\limits_{U_{\alpha_{i_0}}}\omega$. On the other hand, since $\operatorname{supp}\omega\subset U_{\alpha_{i_0}}$, the partition of unity can be so chosen that $\operatorname{supp}\varphi_\alpha\cap\operatorname{supp}\omega=\varnothing$ for $\alpha\neq\alpha_{i_0}$. Then $\varphi_{\alpha_{i_0}}\equiv 1$ on $\operatorname{supp}\omega$, and formula (6) becomes $\int\limits_M\omega=\int\limits_{U_{\alpha_{i_0}}}\varphi_{\alpha_{i_0}}\omega=\int\limits_{U_{\alpha_{i_0}}}\omega$. Statement (a) is proved.

We now prove statement (c). The opposite orientation on the manifold M can be defined without changing the atlas, but changing only the coordinate system in each atlas U_α: namely, $(y_\alpha^1,\ldots,y_\alpha^n)=(-x_\alpha^1,x_\alpha^2,\ldots,x_\alpha^n)$. For the new orientation we have by formula (4)

$$\begin{aligned}\int\limits_{U'}\omega&\\&=\int\underset{U'}{\cdots}\int f'_\alpha(y_\alpha^1,\ldots,y_\alpha^n)\,dy_\alpha^1\ldots dy_\alpha^n\\&=\int\underset{U}{\cdots}\int f'_\alpha(y_\alpha^1(x_\alpha^1,\ldots,x_\alpha^n),\ldots)\left|\det\left(\frac{\partial y_\alpha^i}{\partial x_\alpha^j}\right)\right|dx_\alpha^1\ldots dx_\alpha^n.\end{aligned}$$

Since

$$f_\alpha(x_\alpha^1,\ldots,x_\alpha^n)=f'_\alpha(y_\alpha^1,(x_\alpha^1,\ldots,x_\alpha^n),\ldots)\det\left(\frac{\partial y_\alpha^i}{\partial x_\alpha^j}\right),$$

we have

$$\begin{aligned}\int\limits_{U'}\omega&=\int f_\alpha(x_\alpha^1,\ldots,x_\alpha^n)\operatorname{sgn}\det\left(\frac{\partial y_\alpha^i}{\partial x_\alpha^j}\right)dx_\alpha^1\ldots dx_\alpha^n\\&=-\int f_\alpha(x_\alpha^1,\ldots,x_\alpha^n)\,dx_\alpha^1\ldots dx_\alpha^n=-\int\limits_U\omega.\end{aligned}$$

Hence, the sign of all the terms in (5) will also change if the orientation of M is altered. Proposition 1 is proved.

6.2.2. STOKES' THEOREM

We now prove the fundamental formula which generalizes numerous integration formulas of mathematical analysis.

Theorem 1. *Let M be an n-dimensional oriented manifold with boundary ∂M and let ω be an exterior differential form of degree $(n-1)$ with compact support. Then*

$$(-1)^n \int_M d\omega = \int_{\partial M} \omega. \tag{9}$$

Formula (9) is called *Stokes' formula*. Before turning to the *proof* of Theorem 1, we discuss particular cases of this formula.

Examples

1. Consider in the plane $\mathbf{R}^2$ a smooth closed curve Γ without self-intersections which bounds an open domain V in $\mathbf{R}^2$. Let on Γ there be valid a parameter t defining the circumvention

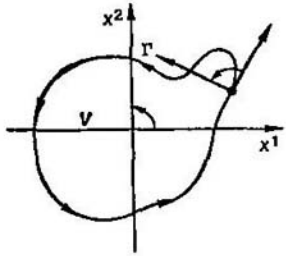

Figure 6.1

direction and, therefore, the orientation of Γ as a one-dimensional manifold. Then the closure $\overline{V}$ is an oriented two-dimensional manifold with the boundary $\partial\overline{V} = \Gamma$. If the orientation of the domain $\overline{V}$ is defined by a linear coordinate system (x^1, x^2), the orientation on the boundary Γ will be compatible with the orientation of $\overline{V}$, provided the domain $\overline{V}$ lies on the left of Γ when Γ is traversed in the direction of increasing parameter t (see Fig. 6.1). Let ω be an arbitrary form of degree 1 on $\mathbf{R}^2$. In the coordinates (x^1, x^2) the form ω is represented as $\omega = P(x^1, x^2)\,dx^1 + Q(x^1, x^2)\,dx^2$. Then the

integral of ω along the curve Γ coincides with the integral of the second kind

$$\int_{\Gamma} \omega = \int_{\Gamma} P\,dx^1 + Q\,dx^2$$
$$= \int_{t_0}^{t_1} \left(P\left(x^1(t), x^2(t)\right) \frac{dx^1}{dt} + Q\left(x^1(t), x^2(t)\right) \frac{dx^2}{dt} \right) dt.$$

By formula (9) we have

$$\int_{\Gamma} \omega = \int_{V} d\omega = \int_{V} (dP \wedge dx^1 + dQ \wedge dx^2)$$
$$= \int_{V} \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1 \wedge dx^2 = \iint_{V} \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1\,dx^2.$$

Thus, we obtain *Green's formula* in two dimensions

$$\int_{\Gamma} (P\,dx^1 + Q\,dx^2) = \iint_{V} \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1\,dx^2.$$

2. Similarly, let Γ be a smooth closed curve without self-intersections in $\mathbf{R}^3$, and let this curve be the boundary of a two-dimensional surface V. By formula (9) we obtain *Stokes' formula* in three dimensions. Let us consider a form ω of degree 1 in a three-dimensional space $\mathbf{R}^3$

$$\omega = P(x^1, x^2, x^3)\,dx^1 + Q(x^1, x^2, x^3)\,dx^2 + R(x^1, x^2, x^3)\,dx^3.$$

Then the integral of the second kind over the curve Γ may be interpreted as the integral of the form ω

$$\int_{\Gamma} \omega = \int_{t_0}^{t_1} \left(P(x^1, x^2, x^3) \frac{dx^1}{dt} + Q(x^1, x^2, x^3) \frac{dx^2}{dt} + R(x^1, x^2, x^3) \frac{dx^3}{dt} \right) dt.$$

From formula (9) we have $\int_{\Gamma} \omega = \int_{V} d\omega$,

$$d\omega = dP \wedge dx^1 + dQ \wedge dx^2 + dR \wedge dx^3$$
$$= \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1 \wedge dx^2 + \left(\frac{\partial R}{\partial x^2} - \frac{\partial Q}{\partial x^3} \right) dx^2 \wedge dx^3$$
$$+ \left(\frac{\partial P}{\partial x^3} - \frac{\partial R}{\partial x^1} \right) dx^3 \wedge dx^1.$$

Thus, the integral along the curve Γ is expressed in terms of the integral of the second kind over the surface V

$$\int_{\Gamma} (P\,dx^1 + Q\,dx^2 + R\,dx^3)$$
$$= \iint_V \left\{ \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2}\right) dx^1\,dx^2 + \left(\frac{\partial R}{\partial x^2} - \frac{\partial Q}{\partial x^3}\right) dx^2\,dx^3 + \left(\frac{\partial P}{x^3} - \frac{\partial R}{\partial x^1}\right) dx^3\,dx^1 \right\}.$$

3. The last relation, the *Gauss-Ostrogradsky formula*, is also a particular case of (9). Let Γ be a closed surface in $\mathbf{R}^3$ which bounds a domain V, and let ω be an exterior differential form of degree 2. In coordinates (x^1, x^2, x^3) ω is represented as

$$\omega = P dx^1 \wedge dx^2 + Q\,dx^2 \wedge dx^3 + R dx^3 \wedge dx^1.$$

Then the integral of the second kind over the surface Γ may be interpreted as the integral of the form ω. Applying formula (9), we obtain

$$\int_{\Gamma} \omega = -\int_{V} d\omega.$$

$$d\omega = dP \wedge dx^1 \wedge dx^2 + dQ \wedge dx^2 \wedge dx^3 + dR \wedge dx^3 \wedge dx^1$$
$$= \left(\frac{\partial P}{\partial x^3} + \frac{\partial Q}{\partial x^1} + \frac{\partial R}{\partial x^2}\right) dx^1 \wedge dx^2 \wedge dx^3,$$

whence

$$\iint_{\Gamma} (P\,dx^1\,dx^2 + Q\,dx^2\,dx^3 + R\,dx^3\,dx^1)$$
$$= -\iiint_V \left(\frac{\partial P}{\partial x^3} + \frac{\partial Q}{\partial x^1} + \frac{\partial R}{\partial x^2}\right) dx^1\,dx^2\,dx^3.$$

Thus, we have arrived at the Gauss-Ostrogradsky formula with the opposite sign. The minus sign appears because in the classical Gauss-Ostrogradsky formula the orientation of the surface Γ, defined by a pair of vectors $(\mathbf{e}_1, \mathbf{e}_2)$ in the tangent space to Γ, is so chosen that the third vector $\mathbf{e}_3$ is directed outward to the domain V.

4. Finally, the *Newton-Leibniz formula* $\int_a^b \frac{df}{dx}(x)\,dx = f(b) - f(a)$ can also be considered as a particular case of formula (9). The left-hand side of the Newton-Leibniz formula may be interpreted as the integral of the form $df = \frac{df}{dx}\,dx$ over the real segment $[a, b]$ which is a one-dimensional oriented manifold with boundary. The

boundary of the segment $[a, b]$ is represented by two points $\{a, b\}$ which we shall consider as a 0-dimensional manifold. Thus far, however, we have not defined the concept of orientation for a 0-dimensional manifold because such a manifold does not have a tangent vector. But we can proceed in the following way. A point of the boundary of a one-dimensional manifold M is assumed to have positive orientation if the vector (at this point) directed inward to M gives the initial orientation of M; on the contrary, this point has negative orientation if the inward vector gives the orientation opposite to that of M. A zero-dimensional form on a manifold is simply a function. Thus, the integral of a function over a zero-dimensional manifold is the sum of the values of the function at points, the sign of the values being taken in accordance with orientation of these points on the manifold. In our case the right-hand side of the Newton-Leibniz formula may be interpreted as the integral of the function f over the boundary consisting only of two points $\{a, b\}$, the point a having positive orientation and the point b negative orientation. Hence, $\int\limits_{\partial[a,b]} f = f(a) - f(b)$, and formula (9) takes the form $\int\limits_{[a,b]} df = (-1)^1 \int\limits_{\partial[a,b]} f$.

Proof of Theorem 1. Since the two sides of formula (9) are linear with respect to the form ω, we can decompose ω into the sum $\omega = \omega_1 + \ldots + \omega_N$ and prove the formula for the case where the support of ω is compact and belongs to one chart. All the more, it is sufficient to prove formula (9) for ω of the form

$$\omega = f(x^1, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^{k-1} \wedge dx^{k+1} \wedge \ldots \wedge dx^n,$$

where f is a function with compact support defined in $\mathbf{R}^n_+$. In this case $d\omega\, (-1)^{k-1} \frac{\partial f}{\partial x^k}\, dx^1 \wedge \ldots \wedge dx^n$. Suppose first that $k < n$. Then the restriction of ω to the boundary $\mathbf{R}^{n-1}_0 \subset \mathbf{R}^n_+$ vanishes because $dx^n = 0$, so that $\int\limits_{\mathbf{R}^{n-1}_0} \omega = 0$. On the other hand, $\int\limits_{\mathbf{R}_+} d\omega = \int \underset{\mathbf{R}^n_+}{\ldots} \int (-1)^{k-1} \frac{\partial f}{\partial x^k}\, dx^1 \ldots dx^n$. Going over to the multiple integral, first with respect to the variable x^k and then with respect to all other variables, we obtain

$$\int\limits_{\mathbf{R}^n_+} d\omega = \int \underset{\mathbf{R}^{n-1}_+}{\ldots} \int dx^1 \ldots dx^{k-1}\, dx^{k+1} \ldots dx^n \int\limits_{-\infty}^{\infty} \frac{\partial f}{\partial x^k}\, dx^k.$$

The interior integral vanishes due to the following obvious relations: $\int\limits_{-\infty}^{\infty} \frac{\partial f}{\partial x^k} dx^k = f(x^1, \ldots, x^n) \big|_{x^k=-\infty}^{x^k=+\infty}$. Thus, $\int\limits_{\mathbf{R}_+^n} d\omega = 0$. Let now $k = n$. Then $\omega = f(x^1, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^{n-1}$,

$$d\omega = (-1)^{n-1} \frac{\partial f}{\partial x^n}(x^1, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^{n-1} \wedge dx^n.$$

We have

$$\int\limits_{\mathbf{R}_0^{n-1}} \omega = \int \underset{\mathbf{R}_0^{n-1}}{\cdots} \int f(x^1, \ldots, x^{n-1}, 0)\, dx^1 \ldots dx^{n-1},$$

$$\int\limits_{\mathbf{R}_+^n} d\omega = \int \underset{\mathbf{R}_+^n}{\cdots} \int (-1)^{n-1} \frac{\partial f}{\partial x^n} dx^1 \ldots dx^{n-1}\, dx^n$$

$$= \int \underset{\mathbf{R}_0^{n-1}}{\cdots} \int dx^1 \ldots dx^{n-1} \int\limits_0^{\infty} (-1)^{n-1} \frac{\partial f}{\partial x^n} dx^n$$

$$= \int \underset{\mathbf{R}_0^{n-1}}{\cdots} \int dx^1 \ldots dx^{n-1} \left((-1)^{n-1} f(x^1 \ldots x^n) \big|_{x^n=0}^{x^n=+\infty}\right)$$

$$= (-1)^n \int \underset{\mathbf{R}_0^{n-1}}{\cdots} \int f(x^1 \ldots x^{n-1}, 0)\, dx^1 \ldots dx^{n-1}.$$

Hence, $\int\limits_{\mathbf{R}_0^{n-1}} \omega = (-1)^n \int\limits_{\mathbf{R}_+^n} d\omega$. To complete the proof, it is sufficient to decompose the form ω into the sum $\omega = \omega_1 + \ldots + \omega_N$ in such a way that the support of each term be contained in one chart of the manifold M. Let us consider an atlas $\{U_\alpha\}$ and a partition of unity $\{\varphi_\alpha\}$ subordinate to the atlas $\{U_\alpha\}$. Then $1 \equiv \Sigma\varphi_\alpha$, $\omega = \Sigma\varphi_\alpha\omega$, $\operatorname{supp} \varphi_\alpha\omega \subset U_\alpha$. Theorem 1 is proved.

Example 5. Theorem 1 can be used to construct a cohomology class defined by a closed form ω on a manifold. We note first of all that *if ω is an exact form*, i.e. it defines a null cohomology class $[\omega] = 0 \in H^k(M)$, *the integral of the form ω over any closed oriented submanifold $W \subset M$*, $\dim W = k$, *vanishes*, $\int\limits_W \omega = 0$. Indeed, since $\omega = d\Omega$ and $\partial W = \varnothing$, we have by Stokes' formula $\int\limits_W \omega = (-1)^{k-1} \int\limits_{\partial W} \Omega = 0$. The converse is also true: *if the integral of a closed form $f^*\omega$ of degree k over any closed oriented manifold W*, $\dim W = k$, *mapped into M, $f\colon W \to M$, vanishes, then the form ω is exact.*

We shall verify this statement for a particular case $k = 1$. In this case, the condition that $\int\limits_{\gamma} \omega = 0$ for any closed curve γ means that integral $\int\limits_{\gamma} \omega$ along an unclosed curve depends only on the initial and terminal points of the curve γ. Then the function f, such that $df = \omega$, is sought in the form $f(P) = \int\limits_{\gamma} \omega$, where γ is a curve connecting a fixed point P_0 with a "variable" point P. To prove the equality $df = \omega$, it is sufficient to choose a local coordinate system in the neighbourhood of P, say $(x^1, \ldots, x^n)$. Then, as γ we can always choose the curve comprising the following trajectories: the fixed curve γ_0 which connects P_0 with the origin $(0, \ldots, 0)$ and the sequence of segments connecting the points $(0, \ldots, 0)$, $(x^1, \ldots, 0)$, $(x^1, x^2, \ldots, 0)$, $\ldots$, $(x^1, \ldots, x^{n-1}, 0)$, $(x^1, \ldots, x^{n-1}, x^n)$. Putting $\omega = \sum g_i \, dx^i$, we obtain

$$f(x^1, \ldots, x^n)$$

$$= \int\limits_{\gamma_0} \omega + \int\limits_0^{x_1} g_1(x^1, \ldots, 0)\, dx^1$$

$$+ \int\limits_0^{x^2} g_2(x^1, x^2, \ldots, 0)\, dx^2 + \ldots + \int\limits_0^{x^n} g_n(x^1, \ldots, x^n)\, dx^n.$$

Hence,

$$\frac{\partial f}{\partial x^k} = g_k(x^1, \ldots, x^k, \ldots 0)$$

$$+ \int\limits_0^{x^{k+1}} \frac{\partial}{\partial x^k} g_{k+1}(x^1, \ldots, x^{k+1}, \ldots, 0)\, dx^{k+1}$$

$$+ \int\limits_0^{x^n} \frac{\partial}{\partial x^n} g_n(x^1, \ldots, x^n)\, dx^n = g_k(x^1, \ldots, x^k, \ldots, 0)$$

$$+ \int\limits_0^{x^{k+1}} \frac{\partial g_k}{\partial x^{k+1}}(x^1, \ldots, x^{k+1})\, dx^{k+1}$$

$$+ \ldots + \int\limits_0^{x^n} \frac{\partial}{\partial x^n} g_k(x^1, \ldots, x^n)\, dx^n = g_k(x^1, \ldots, x^k, \ldots, 0)$$

$$+ [g_k(x^1, \ldots, x^{k+1}, \ldots, 0) - g_k(x^1, \ldots, x^k, \ldots, 0)]$$

$$+ \ldots + [g_k(x^1, \ldots, x^n) - g_k(x^1, \ldots, x^{n-1}, 0)] = g_k(x^1, \ldots, x^n)$$

6.3. THE DEGREE OF MAPPING AND ITS APPLICATIONS

This section deals with an important geometric version of mappings, the degree of a mapping.

6.3.1. EXAMPLE

Let us consider a circle S^1 realized as the set of complex numbers with modulus equal to unity, and a mapping $f\colon S^1 \to S^1$, $f(z) = z^n$. This mapping is smooth. Any point of S^1 is regular for

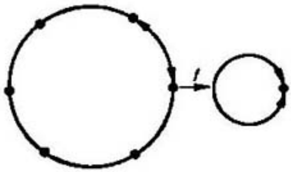

Figure 6.2

the mapping f. Indeed, in the local parameter φ the mapping f is of the form $f(\varphi) = n\varphi$, $df(\xi) = n\xi$, $n \neq 0$. The differential df is thus an isomorphism. The inverse image of any point $z_0 \in S^1$ consists of exactly n points, the roots of order n of the complex number z_0. Geometrically, the mapping f may be looked upon as an n-fold "winding" of S^1 onto S^1 (Fig. 6.2). Let us consider an exterior differential form ω equal, in the local parameter, to $\omega = d\varphi$. Then, $f^*(\omega) = d(f(\varphi)) = d(n\varphi) = n\,d\varphi = n\omega$ and therefore

$$\int_{S^1} f^*(\omega) = n \int_{S^1} \omega. \tag{1}$$

On the other hand, subdividing the circle into n segments $I_1, \ldots, I_n$ (viz., by $\sqrt[n]{1}$), we find that on each segment I_k the mapping f is a diffeomorphism onto S^1 (without one point). Thus, since the integral of an exterior differential form is invariant under coordinate transformations (with positive Jacobian), we obtain $\int_{I_k} f^*(\omega) = \int_{S^1} \omega$. Hence, relation (1) can be derived as follows: we calculate the number of inverse images of a regular point and multiply the integral of the form ω by this number, which gives the left-hand side of (1).

By virtue of Example 5 (Sec. 6.2), we find that on the circle S^1 the behaviour of the homomorphism $f^*\colon H^1(S^1) \to H^1(S^1)$ in cohomology groups is determined by the same number n.

6.3.2. THE DEGREE OF MAPPING

Definition 1. Let $f: M_1 \to M_2$ be a smooth mapping of compact, connected, oriented, closed manifold, $\dim M_1 = \dim M_2$, and let $P \in M_2$ be a regular point. For $Q \in f^{-1}(P)$ we put $\varepsilon(Q) = +1$ if the determinant of the Jacobi matrix of f at the point Q is positive, and $\varepsilon(Q) = -1$ if the determinant is negative. The *degree of the mapping* f (relative to a regular point P) is the number

$$\deg_P f = \sum_{Q \in f^{-1}(P)} \varepsilon(Q). \tag{2}$$

Theorem 1. *Definition* (1) *does not depend*

(a) *on the choice of the regular point* $P \in M_2$,

(b) *on the choice of the mapping* f *in the class of smoothwise homotopic mappings.*

Proof. Item (a) is reduced to item (b). Indeed, if P and P' are two regular points, there exists a continuous family of diffeomorphisms

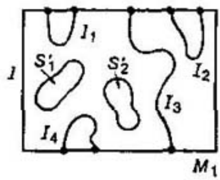

Figure 6.3

$\varphi_t: M_2 \to M_2$ such that $\varphi_0(P) = P$, $\varphi_1(P') = P$. The mappings f and $\varphi_1 \circ f$ are then homotopic and P is a regular point for both of them. On the other hand, $\deg_P(\varphi_1 \circ f) = \deg_{P'} f$. So let $F: M_1 \times I \to M_2$ be a smooth mapping and let $P \in M_2$ be a regular point both for $f_0 = F|_{(M_1 \times \{0\})}$ and for $f_1 = F|_{(M_1 \times \{1\})}$. Then for the mapping F the point P is regular at all points of the boundary $\partial(M_1 \times I) = (M_1 \times \{0\}) \cup (M_1 \times \{1\})$.

According to Sard's theorem, there exists for the mapping F a regular point P' arbitrarily close to P. Then the point P' is also regular for the mappings f_0 and f_1. Since P' can be chosen arbitrarily close to P, for each inverse image $Q \in f_0^{-1}(P)$ there exists a unique close inverse image $Q' \in f_0^{-1}(P')$ such that $\varepsilon(Q) = \varepsilon(Q')$. Hence, $\deg_P f_0 = \deg_{P'} f_0$. By similar reasoning we find that $\deg_P f_1 = \deg_{P'} f_1$. Again, denoting the point P' by P, we obtain that the inverse image $F^{-1}(P)$ is a smooth one-dimensional manifold with the boundary $\partial F^{-1}(P)$ which belongs to $\partial(M_1 \times I)$. A one-dimensional compact manifold is always the union of its connected components, circles and segments (see Fig. 6.3). Hence, the manifold $F^{-1}(P)$

can also be decomposed into the topological sum of segments I_k and circles S_l^1: $F^{-1}(P) = \bigcup_k I_k \cup \bigcup_l S_l^1$. Each segment I_k has two boundary points, a_k and b_k. The set of all points $\{a_k, b_k\}$ forms the union of the inverse images $f_0^{-1}(P)$ and $f_1^{-1}(P)$. We now demonstrate that if the pair (a_k, b_k) belongs to the same component of the boundary $\partial(M_1 \times I)$, then $\varepsilon(a_k) = -\varepsilon(b_k)$, and if this pair belongs to different components, then $\varepsilon(a_k) = \varepsilon(b_k)$. Indeed, as an atlas on the manifold $M_1 \times I$ we can choose the charts $U_\alpha \times I$ with the coordinates $(x_\alpha^1, \ldots, x_\alpha^n, t)$. On the segment I_k we define a parameter φ, $0 \leqslant \varphi \leqslant 1$. First, let the points a_k, b_k lie on the connected component $M_1 \times \{0\}$, $a_k \in U_\alpha \times \{0\}$, $b_k \in U_\beta \times \{0\}$. Then the following inequalities are satisfied:

$$\left.\frac{\partial t}{\partial \varphi}\right|_{a_k} > 0, \quad \left.\frac{\partial t}{\partial \varphi}\right|_{b_k} < 0. \tag{3}$$

On the other hand, for a sufficiently small neighbourhood $V \ni P$ all inverse images from $F^{-1}(V)$ represent, in a neighbourhood of I_k, segments parametrized by the same parameter φ, so that $F^{-1}(V) = V \times F^{-1}(P) = V \times I_k$. If the neighbourhood V is furnished with coordinates $(y^1, \ldots, y^n)$, in the domain $F^{-1}(V)$ we can choose the coordinates $(y^1, \ldots, y^n, \varphi)$. Thus, to calculate the numbers $\varepsilon(a_k)$ and $\varepsilon(b_k)$, it suffices to calculate the sign of the determinant of the matrices $\left(\frac{\partial x_\alpha^i}{\partial y^j}\right)$, $\left(\frac{\partial x_\beta^i}{\partial y^j}\right)$. Since the manifold $M_1 \times I$ is oriented, the determinants of the Jacobi matrices of the coordinate transformation $(x_\alpha^1, \ldots, x_\alpha^n, t) \to (y^1, \ldots, y^n, \varphi)$ have the same sign, irrespective of the index α. For example,

$$\left.\frac{\partial(x_\alpha^1, \ldots, x_\alpha^n, t)}{\partial(y^1, \ldots, y^n, \varphi)}\right|_{a_k} = \frac{\partial(x_\alpha^1, \ldots, x_\alpha^n)}{\partial(y^1, \ldots, y^n)} \cdot \left.\frac{\partial t}{\partial \varphi}\right|_{a_k}. \tag{4}$$

Hence, the first factors on the right-hand side of (4) have different signs for a_k and b_k. If $a_k \in M_1 \times \{0\}$ and $b_k \in M_1 \times \{1\}$, we obtain, instead of (3),

$$\left.\frac{\partial t}{\partial \varphi}\right|_{a_k} > 0, \quad \left.\frac{\partial t}{\partial \varphi}\right|_{b_k} > 0. \tag{5}$$

Formula (4) implies, therefore, that $\varepsilon(a_k) = \varepsilon(b_k)$. Thus, all the points of the inverse images $f_0^{-1}(P)$ and $f_1^{-1}(P)$ split into pairs satisfying the condition: if a pair belongs to the same component of the boundary $\partial(M_1 \times I)$, it gives zero contribution to the sum (2), and if a pair belongs to distinct components of $\partial(M_1 \times I)$, it gives the same contribution to (2) both for $\deg_P f_0$ and for $\deg_P f_1$. Theorem 1 is proved.

Remark. For non-orientable manifolds an analogue of the degree of a mapping can also be defined by formula (2). In this case, instead of Theorem 1 we should assert that the degree of a mapping does not depend on the choice of the point and homotopy with respect to mod 2.

6.3.3. THE FUNDAMENTAL THEOREM OF ALGEBRA

The fundamental theorem of algebra states that *any polynomial $P(z)$ of degree $\geqslant 1$ over the field of complex numbers has at least one complex root.*

There are many various proofs of this theorem. One of them rests on using the concept of the degree of a mapping and Theorem 1. Let us consider a smooth mapping $P\colon \mathbf{C}^1 \to \mathbf{C}^1$ of the complex plane defined by

$$w = P(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0. \tag{6}$$

This mapping can be extended to the mapping of a two-dimensional sphere S^2 into itself, assuming S^2 to be a complex projective straight line $\mathbf{CP}(1)$. To this end, we assume that the complex parameter z is equal to the ratio of homogeneous coordinates on $\mathbf{CP}(1)$: $z = \frac{z_1}{z_0}$ for $z_0 \neq 0$. Similarly, $w = \frac{w_1}{w_0}$ for $w_0 \neq 0$. Therefore, the mapping

$$w_1 = z_1^n + a_{n-1}z_1^{n-1}z_0 + \ldots + a_1 z_1 z_0^{n-1} + a_0 z_0^n, \qquad w_0 = z_0^n \tag{7}$$

correctly defines the mapping of $\mathbf{CP}(1)$ into itself. Mapping (7) is, apparently, a smooth one. Indeed, in the chart $z_0 \neq 0$ this follows from (6), in the chart $z_1 \neq 0$ as a complex coordinate we can take the function $z' = \frac{z_0}{z_1}$. By setting $w' = \frac{w_0}{w_1}$, we obtain

$$w' = (z')^n (1 + a_{n-1}z' + \ldots + a_1 (z')^{n-1} + a_0 (z')^n)^{-1}. \tag{8}$$

Taking a sufficiently small $\varepsilon > 0$, we choose a chart containing the point $z' = 0$ and define this chart by the inequality $|z'| < \varepsilon$ such that the denominator in (8) be non-zero. Thus, the mapping f: $\mathbf{CP}(1) \to \mathbf{CP}(1)$ given by formula (7) is smooth. We now calculate the degree of f. According to Theorem 1, the mapping f can be replaced by a homotopic one. Let us consider the homotopy with respect to the parameter t, $0 \leqslant t \leqslant 1$, defined by

$$w_1 = z_1^n + t(a_{n-1}z^{n-1}z_0 + \ldots + a_0 z_0^n), \qquad w_0 = z_0^n. \tag{9}$$

Just like in the case of (7), mappings (9) are smooth. At $t = 0$ we obtain a simple mapping

$$w_1 = z_1^n, \quad w_0 = z_0^n. \tag{10}$$

In the local coordinates $w = w_1/w_0$, $z = z_1/z_0$ this mapping takes the form $w = z^n$, and, say, the point $w = 1$ is regular. Indeed, calculating the Jacobi matrix of the mapping $u = \operatorname{Re} w = \operatorname{Re} z^n$, $v = \operatorname{Im} w = \operatorname{Im} z^n$, $z = x + iy$, we obtain

$$\det \frac{\partial (u, v)}{\partial (x, y)} = \det \begin{pmatrix} \operatorname{Re} \frac{\partial w}{\partial z} & -\operatorname{Im} \frac{\partial w}{\partial z} \\ \operatorname{Im} \frac{\partial w}{\partial z} & \operatorname{Re} \frac{\partial w}{\partial z} \end{pmatrix}$$
$$= n^2 \det \begin{pmatrix} \operatorname{Re} z^{n-1} & -\operatorname{Im} z^{n-1} \\ \operatorname{Im} z^{n-1} & \operatorname{Re} z^{n-1} \end{pmatrix} = n^2 \mid z^{n-1} \mid^2 > 0$$

for $z \neq 0$. Since the equation $z^n = 1$ has exactly n solutions, the degree of mapping (10) and, therefore, of mapping (7) is equal to n, i.e. $\deg f = n$. If the polynomial P did not have roots, the point $w = 0$ would not belong to the image of f and, hence, the mapping f: $\mathbf{CP}(1) \to \mathbf{CP}(1)$ would have a regular point ($w = 0$) with empty inverse image, i.e. the degree of mapping f would be zero. Contradiction proves the theorem.

6.3.4. INTEGRATION OF FORMS

Theorem 2. *Let f: $M_1 \to M_2$ be a smooth mapping of oriented, compact, connected manifolds, and let ω be an exterior differential form,* $\deg \omega = \dim M_1 = \dim M_2$. *Then*

$$\int_{M_1} f'(\omega) = \deg f \cdot \int_{M_2} \omega. \tag{11}$$

Proof. Since the left-hand and right-hand sides of (11) are both linear in ω, it is sufficient to verify this formula for the form ω whose support lies in a small neighbourhood U of the point $Q \in M_2$. Let $Q_0 \in M_2$ be a regular point of the mapping f and let $U_0 \ni Q_0$ be a sufficiently small neighbourhood. Then there exists a continuous family of diffeomorphisms φ_t: $M_2 \to M_2$ such that $\varphi_0(P) = P$ is an identity mapping and $\varphi_1(U_0) = U \ni Q$. Indeed, connect Q and Q_0 by a continuous path γ. Without loss of generality, we may assume that Q and Q_0 belong to the same chart V diffeomorphic to $\mathbf{R}^n$, and the path γ is a straight segment in the local coordinate system. Let us construct on γ a vector field ξ which is equal to the tangent vector to γ and has a compact support in the chart V. In this case the dynamical system corresponding to the vector field ξ,

i.e. the one-parameter family of diffeomorphisms φ_t, shifts the point Q to the point Q_0, and the diffeomorphism φ_0 is an identity one for $t = 0$. The forms $\varphi_0^*(\omega) = \omega$ and $\varphi_1^*(\omega) = \omega_1$ are, therefore, cohomologous (according to Theorem 4 of Sec. 6.1), and by Stokes' theorem we have $\int\limits_{M_2} \omega = \int\limits_{M_2} \omega_1$. Similarly, $\int\limits_{M_1} f^*(\omega) = \int\limits_{M_1} f^*(\omega_1)$.

Since $\operatorname{supp} \omega \subset U$, we have $\operatorname{supp} \omega_1 \subset U_0$. Hence, $\int\limits_{M_2} \omega_1 = \int\limits_{U_0} \omega_1$.

On the other hand, the inverse image $f^{-1}(U_0)$ is the union of finitely many open sets $f^{-1}(U_0) = \bigcup\limits_{i=1}^{N} V_i$, the mapping f being a diffeomorphism on each of these sets. Thus, since $\operatorname{supp} f^*(\omega_1) \subset f^{-1}(U_0)$, we have

$$\int\limits_{M_1} f^*(\omega_1) = \int\limits_{f^{-1}(U_0)} f^*(\omega_1) = \sum_{i=1}^{N} \int\limits_{V_i} f^*(\omega_1)$$

$$= \sum_{i=1}^{N} (\operatorname{sgn} df\,|_{V_i}) \int\limits_{U_0} \omega_1 = \left[\sum_{i=1}^{N} (\operatorname{sgn} df\,|_{V_i})\right] \int\limits_{U_0} \omega_1$$

$$= \left[\sum_{P \in f^{-1}(Q_0)} \varepsilon(P)\right] \int\limits_{U_0} \omega_1 = \deg f \cdot \int\limits_{U_0} \omega_1.$$

6.3.5. GAUSSIAN MAPPING OF A HYPERSURFACE

Let us consider a hypersurface M in Euclidean space $\mathbf{R}^n$, $\dim M = n - 1$, defined by the equation $F(x) = 0$, $\operatorname{grad} F \neq 0$. Then, there are valid on M the Riemannian metric $\{g_{ij}\}$, the form of volume $\sqrt{|g|}\, dy^1 \wedge \ldots \wedge dy^{n-1}$, where $|g| = \det g_{ij}$, and the Gaussian curvature $K = K(y^1, \ldots, y^n)$. Let us define a new form $K\, d\sigma = K(y^1, \ldots, y^n) \sqrt{|g|}\, dy^1 \wedge \ldots \wedge dy^{n-1}$ called the *form of curvature* of hypersurface M. Furthermore, there is valid on M a smooth mapping $f\colon M \to S^{n-1} \subset \mathbf{R}^n$ associating with each point P the normal to the surface at this point. This mapping is called *spherical mapping*. Let Ω be a form of volume on the sphere S^{n-1}. Then the following statement holds true.

Theorem 3. *For the spherical mapping $f\colon M \to S^{n-1}$ of a hypersurface M the inverse image of the form of volume Ω on the sphere S^{n-1} is equal to the form of curvature on M, $f^*\Omega = K\, d\sigma$.*

Proof. Without loss of generality, we may assume that in the neighbourhood of a point $P_0 \in M$ the manifold M is the graph of the function $x^n = f(x^1, \ldots, x^{n-1})$ and at $P_0 = (0, \ldots, 0)$ the normal to M is parallel to the axis Ox^n. Then at the point P_0 the Riemannian

metric coincides with an identity matrix, $g_{ij} = \delta_{ij}$, and the Gaussian curvature at P_0 is $K = \det\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)$. On the sphere $S^{n-1} \subset \mathbf{R}^n$, in the neighbourhood of the point $f(P_0) = (0, \ldots, 0, 1)$, we choose the coordinates $(x^1, \ldots, x^{n-1})$. Then the metric on S^{n-1} at $f(P_0)$ is also of diagonal form, $g_{ij} = \delta_{ij}$, and therefore the form Ω on the sphere S^{n-1} at $f(P_0)$ is equal to: $\Omega = dx^1 \wedge \ldots \wedge dx^{n-1}$. We now calculate the spherical mapping. The tangent space to the manifold M is generated by the tangent vectors

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ f_{x^1} \end{pmatrix}, \quad \mathbf{r}_2 = \begin{pmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ f_{x^2} \end{pmatrix}, \quad \ldots, \quad \mathbf{r}_{n-1} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ f_{x^{n-1}} \end{pmatrix}.$$

Then the normal vector $\mathbf{n}$ has the coordinates

$$\mathbf{n} = \frac{1}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}} \begin{pmatrix} f_{x^1} \\ f_{x^2} \\ \cdot \\ \cdot \\ \cdot \\ f_{x^{n-1}} \\ -1 \end{pmatrix}.$$

The inverse image of the form Ω under spherical mapping is calculated by substituting the differential of the coordinate of the normal vector $\mathbf{n}$ for dx^i:

$$f^*(\Omega) = d\frac{f_{x^1}}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}} \wedge \ldots \wedge d\frac{f_{x^{n-1}}}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}}.$$

Thus, we have

$$d\frac{f_{x^i}}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}} = \frac{df_{x^i}}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}} - \frac{f_{x^i}(f_{x^1}\, df_{x^1} + \ldots + f_{x^{n-1}}\, df_{x^{n-1}})}{\sqrt{(1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2)^3}}.$$

Since at the point P_0 the first partial derivatives of the function f vanish, $f_{x^i}(P_0) = 0$, we have

$$d\frac{f_{x^i}}{\sqrt{1 + f_{x^1}^2 + \ldots + f_{x^{n-1}}^2}} = \sum_{j=1}^{n-1} \frac{\partial^2 f}{\partial x^i\, \partial x^j}\, dx^j$$

at P_0. Hence,

$$f^*(\Omega) = \left(\sum_{j=1}^{n-1} \frac{\partial^2 f}{\partial x^1 \partial x^j} dx^j\right) \wedge \ldots \wedge \left(\sum_{j=1}^{n-1} \frac{\partial^2 f}{\partial x^{n-1} \partial x^j} dx^j\right)$$

$$= \det\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right) dx^1 \wedge \ldots \wedge dx^{n-1} = K\, d\sigma.$$

Theorem 3 is proved.

As a corollary of this theorem, we arrive at the *Gauss-Bonnet theorem.*

Theorem 4. *Let M be a closed, compact surface in $\mathbf{R}^3$. Then* $\int_M K\, d\sigma =$ $4\pi\lambda$, *where λ is an integer.*

The theorem can be *proved* by using Theorems 2 and 3:

$$\int_M K\, d\sigma = \int_M f^*(\Omega) = (\deg f) \int_{S^2} \Omega = 4\pi \cdot \deg f.$$

Remark. The Gauss-Bonnet theorem is even more general than Theorem 4. It turns out that the integer λ does not depend on the Riemannian metric and is always equal to $1 - g$, where g is the number of handles on the oriented surface M.

Chapter 7

Simple Variational Problems in Riemannian Geometry

7.1. FUNCTIONAL. EXTREMAL FUNCTIONS. EULER'S EQUATIONS

Variational problems constitute one of the most important classes of mathematical problems closely related to such fundamental physical and mechanical phenomena as motion and stability. For example, we shall see below that geodesic trajectories are solutions of the corresponding variational problem.

Before turning to the general concept of the variation of a functional, we shall say a few words about the functional itself. We are already familiar with the concept of a function $y = f(x)$, where y is a real number and the argument x can be written as the set of numbers $(x^1, \ldots, x^n)$ in a curvilinear coordinate system on a smooth manifold. But not every physical correspondence can be expressed in such a simple form. For instance, we have already considered the following correspondence: with each finite segment of a smooth curve $\gamma(t)$ there is associated its length $l_a^b(\gamma) = \int\limits_a^b |\dot{\gamma}|\,dt$. The correspondence $\gamma(t) \to l_a^b(\gamma(t))$ is not a "function" in the ordinary sense, for in this case an arbitrary smooth curve acts as the "argument". The correspondence $\gamma \to l(\gamma)$ is an important example of a non-linear functional defined on the space of smooth curves $\gamma(t)$. We may extend this example to include some other important correspondences.

Let us consider in $\mathbf{R}^n$ a bounded domain D with smooth boundary ∂D and let $x^1, \ldots, x^n$ be Cartesian coordinates. Consider on D all possible smooth vector functions $\mathbf{f}(x^1, \ldots, x^n) = \mathbf{f}(x^\alpha) = (f^1(x^\alpha), \ldots, f^k(x^\alpha)) = \{f^i(x^\alpha)\}$, where the numbers k and n are independent. The domain D is called the range of the parameters $x^1, \ldots, x^n$. Given two functions $\{f^i(x^\alpha)\}$ and $\{g^i(x^\alpha)\}$, $1 \leqslant i \leqslant k$, then for arbitrary real numbers a and b there exists a new vector function $a\mathbf{f} + b\mathbf{g} = \{af^i(x^\alpha) + bg^i(x^\alpha)\}$, i.e. all smooth vector functions on D form a linear space F. This space is infinite-dimensional. Various functionals will be studied on the space F, and "points" of F (i.e. vector functions) are just the "arguments" of functionals $J[\mathbf{f}]$,

$\mathbf{f} \in F$. But in the simple examples below, the existence of a linear structure in F is of no significance, for we shall not use the operation of addition of vector functions.

While studying functionals, it is useful to recall an analogy with ordinary functions.

Definition. A *functional* J defined on a space F (or on a subset of F) is a continuous mapping of F (or of its subset) into real numbers, $J: F \to \mathbf{R}^1$, and the mapping is not assumed to be linear. If $J[a\mathbf{f} + b\mathbf{g}] = aJ[\mathbf{f}] + bJ[\mathbf{g}]$ (i.e. J is linear), the functional J is called *linear*.

Example. Suppose D is a segment on the real straight line $x^1 = t$ and $\mathbf{f}(t) = (f^1(t), f^2(t), f^3(t))$ is a vector function on D, i.e. $\mathbf{f}(t) = \gamma(t)$ defines a smooth curve in $\mathbf{R}^3$; in this case F is a linear space of all such curves in $\mathbf{R}^3$ (the radius vectors $\mathbf{f}$ and $\mathbf{g}$ of the curves $\mathbf{f}(t)$ and $\mathbf{g}(t)$ can be added and multiplied by a number). The functional J is taken as the integral $J[\mathbf{f}] = \int_0^1 |\dot{\gamma}(t)|\,dt = \int_D |\mathbf{f}(t)|\,dt$, i.e. the length of the curve $\mathbf{f}(t)$, $0 \leqslant t \leqslant 1$. This functional is non-linear, since $J[a\mathbf{f} + b\mathbf{g}] \neq aJ[\mathbf{f}] + bJ[\mathbf{g}]$ (give an example!).

Given a smooth function $L(x^\beta; p^i; q^i_\alpha)$ which depends on three groups of variables: x^β, $1 \leqslant \beta \leqslant n$, p^i, $1 \leqslant i \leqslant k$, and q^i_α, $1 \leqslant \alpha \leqslant n$, $1 \leqslant i \leqslant k$. This function is called a *Lagrangian*. Thus, any smooth function of three groups of variables can be a Lagrangian. Let $\mathbf{f} = \{f^i(x^\beta)\}$ be a smooth vector function on $D \subset \mathbf{R}^n$. Let us construct the functional $J[\mathbf{f}]$

$$J[\mathbf{f}] = \int_D L(x^\beta, f^i(x^\beta), f^i_{x^\alpha}(x^\beta))\,d\sigma^n,$$

where $\int_D$ denotes a multiple integral $\int \underset{D}{\cdots} \int$ (n times) over n-dimensional domain D, $d\sigma^n = dx^1 \wedge \ldots \wedge dx^n$ is an n-dimensional elementary volume in D (i.e. a Cartesian exterior form of Euclidean volume), $f^i_{x^\alpha}(x^\beta) = \dfrac{\partial f^i(x^\beta)}{\partial x^\alpha}$ are partial derivatives. The functional $J[\mathbf{f}]$ can be written in compact form as $J[\mathbf{f}] = \int_D L(x^\beta, f^i, f^i_{x^\alpha})\,d\sigma^n$, where the arguments of f^i and $f^i_{x^\alpha}$ are omitted. The function L (Lagrangian) is therefore defined for each functional. The class of functionals just defined includes virtually all important examples of functionals encountered in mechanics, physics, and in their applications.

Let us consider the functional of the arc length

$$J[\mathbf{f}] = \int_0^1 |\dot{\gamma}(t)|\, dt = \int_0^1 \sqrt{g_{ij}(\gamma(t)) \frac{dy^i(t)}{dt} \frac{dy^j(t)}{dt}}\, dt,$$

i.e. $D = I = [0, 1]$, $0 \leqslant t \leqslant 1$, $n = 1$, $\mathbf{f}(t) = \gamma(t) = (y^1(t), \ldots, y^k(t))$, $\gamma(t)$ is a smooth curve in a k-dimensional space with the Riemannian metric $g_{ij}(y^1, \ldots, y^k)$, and the Lagrangian is of the form

$$L(x^\beta, \mathbf{f}, \mathbf{f}_{x^\alpha}) = L\left(t, y^1, \ldots, y^k, \frac{dy^1}{dt}, \ldots, \frac{dy^k}{dt}\right)$$
$$= \sqrt{g_{ij}(y^1, \ldots, y^k)\, \dot{y}^i \dot{y}^j}.$$

If the curve $\gamma(t)$ on the plane $\mathbf{R}^2$ is defined explicitly, $y = f(x)$, then

$$L(x, f, f_x) = L(f_x) = \sqrt{1 + f_x^2}.$$

This functional is also defined on the curves lying in a smooth manifold M^k. We shall usually consider a small neighbourhood of a particular curve and assume that local coordinates $y^1, \ldots, y^k$ are valid in this neighbourhood. Thus, we shall deal with a k-dimensional Euclidean space provided with a Riemannian (generally, non-Euclidean) metric, and in this case the addition of vector functions defining different curves can only be performed in some neighbourhood of a fixed curve.

Another example: the functional of area $J[\mathbf{f}] = \iint\limits_D \sqrt{EG - F^2}\, dx\, dy$.

Here $D(x, y)$ is the range of the parameters (x, y), and $\mathbf{f} = (u^1(x, y), u^2(x, y), u^3(x, y))$ is a two-dimensional surface in $\mathbf{R}^3$ with the induced metric $ds^2 = E\, dx^2 = 2F\, dx\, dy + G\, dy^2$,

$$L = L(\mathbf{f}_x, \mathbf{f}_y) = \sqrt{EG - F^2} = \sqrt{(\mathbf{f}_x, \mathbf{f}_x)(\mathbf{f}_y, \mathbf{f}_y) - (\mathbf{f}_x, \mathbf{f}_y)^2}.$$

The space F is the linear space of all vector functions $(u^1(x, y), u^2(x, y), u^3(x, y))$ defined on D. We could consider another form of the functional of area: $J[\mathbf{f}] = \iint\limits_D \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy$, where D is a domain on $\mathbf{R}^2(x, y)$, and $\mathbf{f}(x, y) = (x, y, z(x, y))$, i.e. the vector function $\mathbf{f}$ is given explicitly by the graph $z = f(x, y)$ over the domain D in $\mathbf{R}^2 \subset \mathbf{R}^3$; $L = L(z_x, z_y) = \sqrt{1 + z_x^2 + z_y^2}$. Addition of such vector functions means addition of their graphs in $\mathbf{R}^3$ over D. But not every surface in $\mathbf{R}^3$ can be described by the graph of a single-valued function.

What problems are of primary interest in the study of $J[\mathbf{f}]$? Let us discuss analogies with ordinary functions. Consider, for example, functions of one or two variables, $\alpha(t)$ and $\alpha(u, v)$. The behaviour of such functions is largely determined by the number and location of the points t_0 (or (u_0, v_0)) at which $\alpha'(t_0) = 0$ (or $\alpha_u(u_0, v_0) = \alpha_v(u_0, v_0) = 0$), i.e. rad $\alpha = 0$. The points at which grad $\alpha = 0$

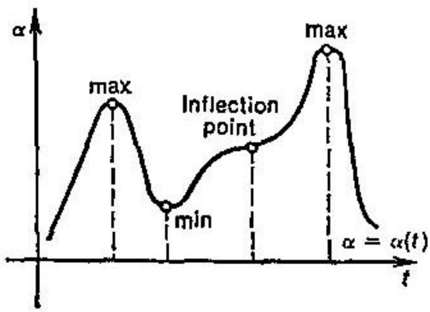

Figure 7.1

are called *critical* or *stationary points* of the function α. The term *extremal points* is also used sometimes. For instance, for the function $\alpha = \alpha(t)$ shown in Fig. 7.1 the stationary points are: two maxima, minimum, and inflection. If $\alpha = \alpha(u, v)$, the points at which grad $\alpha = 0$ are maxima, minima, and saddle points (or simply saddles) of

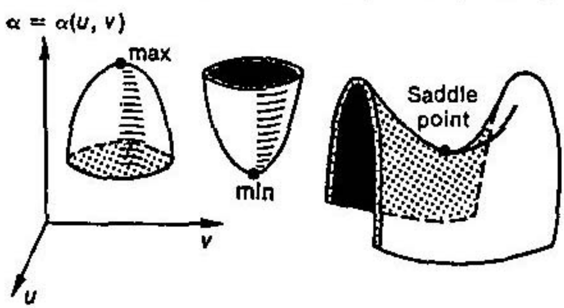

Figure 7.2

order two; saddles of higher orders are also possible (Fig. 7.2). It is of primary importance in mechanics to find the points at which a smooth function, the potential, reaches a maximum (correspondingly, potential energy reaches a minimum), because at these points the system is in stable equilibrium.

Similarly, in the study of $J[\mathbf{f}]$ it is very important to find those stationary vector functions $\mathbf{f}_0$ for which the functional J reaches a

minimum, a maximum, or has a saddle. It is desirable, however, to formulate this geometric analogy in the "differential language", that is, we have to give a correct definition of the directional derivative of a functional J at a point $\mathbf{f} \in F$. As was already noted, all critical points of the functions $\alpha(t)$ and $\alpha(u, v)$ are represented by the solutions of the equation grad $\alpha = 0$. We need, therefore, to derive an analogue of this equation for a functional. Let us consider again the equation grad $\alpha = 0$. If a direction (a vector) $\mathbf{a} = (a^1, a^2)$ is defined at a point $(u, v) \in G$, the derivative of the function $\alpha(u, v)$ with respect to the direction $\mathbf{a}$ is given by

$$\frac{d}{d\mathbf{a}}\alpha(u, v) = a^1\frac{\partial\alpha}{\partial u} + a^2\frac{\partial\alpha}{\partial v} = (\mathbf{a}, \operatorname{grad}\alpha)$$

(see Fig. 7.3). It follows that grad $\alpha(u_0, v_0) = 0$ if and only if $\frac{d}{d\mathbf{a}}\alpha(u_0, v_0) = 0$ for any direction $\mathbf{a}$ at the point (u_0, v_0). If

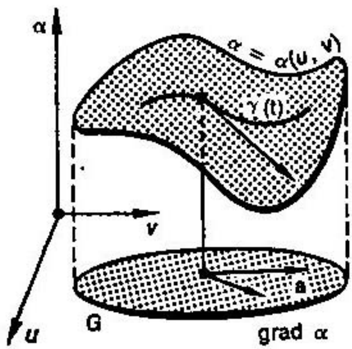

Figure 7.3

$\alpha = \alpha(t)$, this means that $\alpha_t'(t_0) = 0$. The derivative of $\alpha(u, v)$ with respect to $\mathbf{a}$ can be calculated as

$$\frac{d\alpha}{d\mathbf{a}} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}[\alpha(u + \varepsilon a^1, v + \varepsilon a^2) - \alpha(u, v)],$$

where $\mathbf{a} = (a^1, a^2)$ and ε is a parameter, i.e.

$$\frac{d\alpha}{d\mathbf{a}} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}[\alpha(\mathbf{x} + \varepsilon\mathbf{a}) - \alpha(\mathbf{x})],$$

where $\mathbf{x} = (u, v)$ is the radius vector of a point in the domain G. This expression is used to extend the concept of directional derivative to a functional $J[\mathbf{f}]$.

Let us consider a "point" $\mathbf{f} \in F$ and a sufficiently small function $\eta \in F$ such that $\eta|_{\partial D} \equiv 0$. The functions η are called *perturbations* of the function $\mathbf{f}$. Let us consider the shift from the point $\mathbf{f}$ to a point $\mathbf{f} + \varepsilon\eta$ (see Fig. 7.4). The function η (we recall that $\eta|_{\partial D} = 0$)

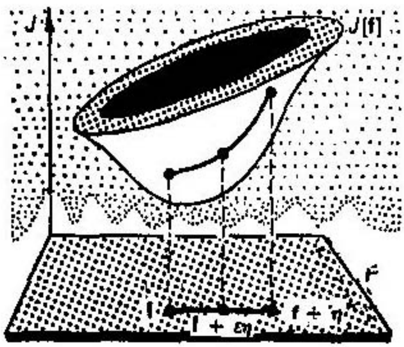

Figure 7.4

defines the "direction of translation" from the "point" $\mathbf{f}$ exactly in the same way as the vector $\mathbf{a}$ defines the direction of translation from the point $(u_0, v_0) \in G$, the difference from an ordinary function being that in the case of functionals there are infinitely many such "directions".

Further, quite similarly to the case of ordinary functions, we may construct the expression $\frac{1}{\varepsilon}(J[\mathbf{f} + \varepsilon\eta] - J[\mathbf{f}])$. Performing a limiting process with respect to ε, we obtain the number

$$\frac{d}{d\eta} J[\mathbf{f}] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(J[\mathbf{f} + \varepsilon\eta] - J[\mathbf{f}])$$

which can naturally be called the *derivative of a functional J at a point* $\mathbf{f}$ *with respect to the direction* η. Drawing further analogy between functionals and ordinary functions, we give another definition.

Definition. A function $\mathbf{f}_0 \in F$ is called a *stationary* (*extremal*, *critical*) function for the functional $J[\mathbf{f}]$, provided $\frac{d}{d\eta} J[\mathbf{f}_0] = 0$ for any perturbation η such that $\eta|_{\partial D} = 0$.

If by F we mean the space of vector functions $\mathbf{f}$ with constant coordinates (i.e. $f^i(x^\alpha) = (\text{const})^i$), these definitions turn into ordinary definitions of directional derivative and stationary point.

It is convenient to represent a functional J as a "graph" over F (Fig. 7.5). Apparently, if $\mathbf{f} \in F$, the set of functions $\mathbf{f} + \varepsilon\eta$, where $\eta|_{\partial D} = 0$, forms a linear space T (provided $\mathbf{f}$ is taken as zero of T);

the functional J is restricted to this space. From this descriptive point of view the "points" $\mathbf{f}_0$, where $\frac{d}{d\eta} J[\mathbf{f}_0] = 0$ (for any η) are minima, maxima, or saddle points of the graph of $J[\mathbf{f}]$ restricted to the subspace $T \subset F$. The meaning of such a restriction of J to T is quite clear: we want to study the behaviour of J for those perturba-

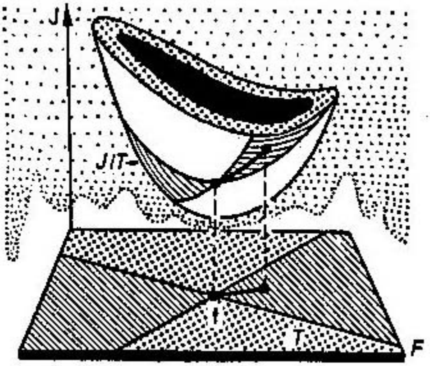

Figure 7.5

tions η which do not alter $\mathbf{f}$ at the boundary ∂D, that is, we study local, differential properties of the function $\mathbf{f}_0$ such that $\frac{d}{d\eta} J[\mathbf{f}_0] = 0$ (for any η). In Fig. 7.6 the ends of the curve are fixed at points A and B,

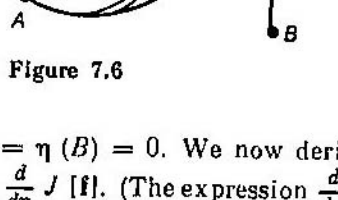

Figure 7.6

i.e. $\eta(A) = \eta(B) = 0$. We now derive an explicit formula for the derivative $\frac{d}{d\eta} J[\mathbf{f}]$. (The expression $\frac{d}{d\eta} J[\mathbf{f}]$ is sometimes called the *first variation* of $J[\mathbf{f}]$.) We have

$$\frac{d}{d\eta} J[\mathbf{f}] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} (J[\mathbf{f} + \varepsilon\eta] - J[\mathbf{f}]).$$

Put $\delta J = J[\mathbf{f}+\varepsilon\eta] - J[\mathbf{f}]$, then

$$\delta J = \int_D [L(x^\beta, f^i + \varepsilon\eta^i, f^i_{x^\alpha} + \varepsilon\eta^i_{x^\alpha}) - L(x^\beta, f^i, f^i_{x^\alpha})]\, d\sigma^n.$$

Representing the integrand as a Taylor series, we obtain

$$\delta J = \int_D \Big[\sum_{i=1}^{k} \frac{\partial L}{\partial f^i}\varepsilon\eta^i + \sum_{i=1}^{k}\sum_{\alpha=1}^{n} \frac{\partial L}{\partial f^i_{x^\alpha}}\varepsilon\eta^i_{x^\alpha} + o(\varepsilon)\Big]\, d\sigma^n$$

$$= \varepsilon \int_D \Big\{\sum_{i=1}^{k}\Big[\frac{\partial L}{\partial f^i}\eta^i + \sum_{\alpha=1}^{n}\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i_{x^\alpha}\Big] + \frac{o(\varepsilon)}{\varepsilon}\Big\}\, d\sigma^n.$$

To integrate by parts, we consider

$$\frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i\right) = \frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)\eta^i + \frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i_{x^\alpha},$$

whence

$$\delta J = \varepsilon \int_D \sum_{\alpha=1}^{n} \frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i\right) d\sigma^n$$

$$+ \varepsilon \int_D \sum_{i=1}^{k}\left(\frac{\partial L}{\partial f^i} - \sum_{\alpha=1}^{n}\frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)\right)\eta^i\, d\sigma^n + \int_D o(\varepsilon)\, d\sigma^n.$$

Since all the functions are assumed smooth, in the first integral we can separate integration with respect to x^α from integration with respect to the other variables x^i $(1 \leqslant i \leqslant n,\ i \neq \alpha)$ (according to the theorem on the change of order of integration). This yields

$$\int_D \frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i\right) d\sigma^n = \int_{x^1 \ldots \hat{x}^\alpha \ldots x^n}\!\!\cdots\int \left[\int_P^Q \frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i\right) dx^\alpha\right] d\sigma^{n-1}$$

(see Fig. 7.7). Since in the integral $\int_P^Q$ the variables $x^1, \ldots, \hat{x}^\alpha, \ldots, x^n$ (x^α is omitted) may be assumed to be parameters, integration can be performed explicitly, i.e.

$$\int_D \frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\eta^i\right) d\sigma^n$$

$$= \int_{x^1 \ldots \hat{x}^\alpha \ldots x^n}\!\!\cdots\int \left[\frac{\partial L}{\partial f^i_{x^\alpha}}(Q)\,\eta^i(Q) - \frac{\partial L}{\partial f^i_{x^\alpha}}(P)\,\eta^i(P)\right] d\sigma^{n-1} \equiv 0,$$

because $\eta^i(P)=\eta^i(Q)=0$, P, $Q\in\partial D$. Thus,

$$\delta J=\varepsilon\int\limits_D\sum_{i=1}^{k}\left[\frac{\partial L}{\partial f^i}-\sum_{\alpha=1}^{n}\frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)\right]\eta^i\,d\sigma^n+\int\limits_D o(\varepsilon)\,d\sigma^n,$$

whence

$$\frac{d}{d\eta}J[\mathbf{f}]=\int\limits_D\sum_{i=1}^{k}\left[\frac{\partial L}{\partial f^i}-\sum_{\alpha=1}^{n}\frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)\right]\eta^i\,d\sigma^n,$$

since $\lim\limits_{\varepsilon\to 0}\frac{1}{\varepsilon}\int\limits_D o(\varepsilon)\,d\sigma^n=0$. Let $\mathbf{f}_0$ be a stationary function for J.

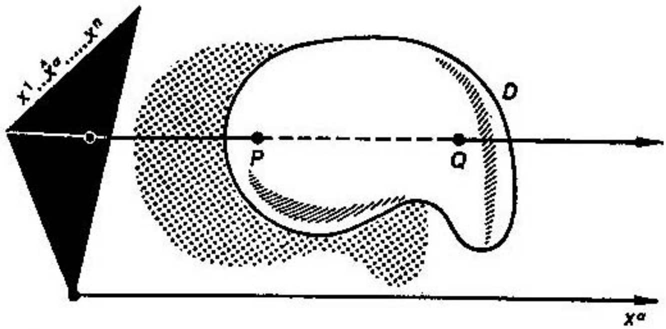

Figure 7.7

Then for any function η ($\eta|_{\partial D}=0$) the following identity is satisfied:

$$\int\limits_D\sum_{i=1}^{k}\eta^i\left[\frac{\partial L}{\partial f^i}-\sum_{\alpha=1}^{n}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)\right]d\sigma^n\equiv 0.$$

As is known from mathematical analysis, this identity means that

$$\frac{\partial L}{\partial f^i}-\sum_{\alpha=1}^{n}\frac{\partial}{\partial x^\alpha}\left(\frac{\partial L}{\partial f^i_{x^\alpha}}\right)=0. \tag{1}$$

System of differential equations (1) is called the *system of Euler's equations for a functional* J [f]. Thus, we have proved the following important statement.

Theorem 1. *A function $f_0\in F$ it called extremal (stationary) function for the functional J [f] if and only if it satisfies Euler's equations* (1).

If a functional is an ordinary function on a domain G, the condition that a point $x_0 \in G$ is extremal means that $\frac{\partial L}{\partial x^i} = 0$, i.e. grad $L = 0$, which is what was to be expected. Here $J = cL$, where $c = \text{const}$.

7.2. EXTREMALITY OF GEODESICS

For a Riemannian manifold a geodesic was defined as a trajectory along which translation preserves the velocity field of the trajectory. But geodesics exhibit another very important characteristic which can be used to define a geodesic. This characteristic is related to the extremal properties of a special functional which is very much alike the functional of length; here geodesics act as extremal solutions to this functional.

Let M^n be a Riemannian manifold with the metric g_{ij} and let $x^1, \ldots, x^n$ be local coordinates. Then the trajectory $\gamma(t)$ can be defined as $\gamma(t) = (x^1(t), \ldots, x^n(t))$; the segment $I = [0, 1]$ is chosen as the domain D. For convenience, we shall consider trajectories $\gamma(t)$ with fixed initial and terminal points: $\gamma(0) = P$, $\gamma(1) = Q$, $P, Q \in M^n$.

Definition. The functional

$$L(\gamma) = \int_0^1 |\dot{\gamma}| \, dt_1 = \int_0^1 \sqrt{g_{ij}(x)\, \dot{x}^i \dot{x}^j}\, dt$$

is called the *functional of the length of the trajectory* $\gamma(t)$. The functional

$$E(\gamma) = \int_0^1 |\dot{\gamma}|^2 \, dt = \int_0^1 g_{ij}(x)\, \dot{x}^i \dot{x}^j \, dt$$

is called the *functional of action of the trajectory* $\gamma(t)$. The functionals L and E are distinct, but they are related by the inequality $L^2 \leqslant E$; there is also a relation between their extremals (see below).

Lemma 1. *The inequality* $L^2 \leqslant E$ *it valid.*

Proof. Applying Schwarz' inequality

$$\left(\int_0^1 fg\, dt\right)^2 \leqslant \left(\int_0^1 f^2\, dt\right) \cdot \left(\int_0^1 g^2\, dt\right)$$

to the functions $f(t) \equiv 1$ and $g(t) = |\dot{\gamma}(t)|$, we obtain $(L(\gamma))^2 \leqslant E(\gamma)$, the equality taking place only for a constant function $g(t)$, i.e. if and only if the parameter t is proportional to the arc length.

Let us consider extremals of the functionals E and L.

Theorem 1. *The extremals of the functional $E(\gamma)$ are represented by geodesics $\gamma(t)$ parametrized by the parameter t proportional to the arc length s. In particular, if the initial condition $|\dot{\gamma}(0)| = 1$ is satisfied at the starting point P, the parameter t is defined uniquely and is the natural parameter, i.e. coincides with the arc length.*

Proof. Recall that we always consider parametrized trajectories. This means that two trajectories describing the same geometric locus and having different parameters are taken as distinct trajectories. According to Theorem 1 of Sec. 6.1, the extremals of the functional $E(\gamma)$ satisfy Euler's equations, which in this case take the form

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0, \quad 1 \leqslant k \leqslant n,$$

and the Lagrangian L is equal to $L(x^i, \dot{x}^j) = g_{ij}(x)\dot{x}^i\dot{x}^j$. Calculation yields

$$\frac{\partial L}{\partial x^k} = \frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j, \qquad \frac{\partial L}{\partial \dot{x}^k} = 2g_{kj}\dot{x}^j,$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 2\frac{\partial g_{kj}}{\partial x^\alpha}\dot{x}^\alpha\dot{x}^j + 2g_{kj}\ddot{x}^j,$$

$$\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j - 2\frac{\partial g_{kj}}{\partial x^i}\dot{x}^i\dot{x}^j - 2g_{kj}\ddot{x}^j = 0,$$

$$g_{kj}\ddot{x}^j + \frac{1}{2}\left(2\frac{\partial g_{kj}}{\partial x^i}\dot{x}^i\dot{x}^j - \frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j\right) = 0,$$

$$\ddot{x}^\alpha + \frac{1}{2}g^{k\alpha}\left(\frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right)\dot{x}^i\dot{x}^j = 0,$$

i.e. $\ddot{x}^\alpha + \Gamma^\alpha_{ij}\dot{x}^i\dot{x}^j = 0$. Thus, Euler's equations coincide with the equations of geodesics in the Riemannian connection. The theorem is proved.

The extremal of the functional $E(\gamma)$ is defined by Euler's equations as a parametrized trajectory. Under an arbitrary smooth change of the parameter on a geodesic trajectory, the trajectory fails, in general, to remain a geodesic.

Theorem 2. *The extremals of the functional $L(\gamma)$ are represented by smooth trajectories $\gamma(t)$ obtained from the geodesic trajectories by arbitrary smooth parameter substitutions. In particular, any extremal of the functional $E(\gamma)$ (a parametrized geodesic) is an extremal of $L(\gamma)$, but the converse does not hold.*

Roughly speaking, the functional $L(\gamma)$ has "more" extremals than the functional $E(\gamma)$.

Proof. Let us consider Euler's equations for $L(\gamma)$. The Lagrangian L is of the form $L(x, \dot{x}) = \sqrt{g_{ij}\dot{x}^i\dot{x}^j}$. We obtain

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = 0, \quad 1 \leqslant i \leqslant n,$$

$$\frac{1}{\sqrt{g_{ij}\dot{x}^i\dot{x}^j}} \frac{\partial}{\partial x^k}(g_{ij}\dot{x}^i\dot{x}^j) - \frac{d}{dt}\left(\frac{1}{\sqrt{g_{ij}\dot{x}^i\dot{x}^j}} \frac{\partial}{\partial \dot{x}^k}(g_{ij}\dot{x}^i\dot{x}^j)\right) = 0.$$

Let $\gamma(t)$ be a solution of this system. Since $\gamma(t)$ is a smooth curve in M^n, this curve admits the natural parameter $t = s$. Then $|\dot{\gamma}(s)| = 1$ along γ and, therefore, Euler's equations become

$$\frac{\partial}{\partial x^k}(g_{ij}\dot{x}^i\dot{x}^j) - \frac{d}{dt}\left(\frac{\partial}{\partial \dot{x}^k}(g_{ij}\dot{x}^i\dot{x}^j)\right) = 0,$$

because $\sqrt{g_{ij}\dot{x}^i\dot{x}^j} = 1$. But these equations coincide, according to Theorem 1, with the equations of geodesics. Thus, if a natural parameter is introduced on an arbitrary solution $\gamma(t)$, an extremal of $L(\gamma)$, this trajectory becomes a geodesic. Conversely, let $\gamma(s)$ be an arbitrary geodesic on M^n; according to the preceding consideration, it is an extremal of $L(\gamma)$. Let $s = s(t)$ be an arbitrary smooth parameter substitution on $\gamma(s)$. Then $L(\gamma(s)) = L(\gamma(s(t)))$, since the length of arc does not change under smooth parameter substitution; hence, the value of L does not change either. Thus, $\gamma(s(t))$ is again an extremal. The theorem is proved.

It is convenient to interpret the relationship between the extremals of E and L as follows. Let us consider the space ΩM^n of all smooth curves on M^n and the action, on this space, of the infinite-dimensional group $\mathfrak{G}$ whose elements are represented by arbitrary smooth parameter substitutions on a curve. Then each point of this space, i.e. a curve γ, generates the orbit $\mathfrak{G}(\gamma)$ of the action of $\mathfrak{G}$ on Ω. Since L is invariant under the action of $\mathfrak{G}$, this functional is constant along each orbit $\mathfrak{G}(\gamma)$, $\gamma \in \Omega$ (Fig. 7.8). Also, since $L(\gamma)$ is constant along each orbit, all points on the orbit $\mathfrak{G}(\gamma)$ are degenerate in the sense that there exist curves (i.e. parametrized trajectories) arbitrarily close to the curve γ, on which L has the same value as on γ. The situation is different, however, for E. This functional varies under parameter substitution and is therefore not constant on the orbits of the action of $\mathfrak{G}$. Thus, to find all the extremals of L, one should consider the orbits of all the extremals of the functional E.

Geodesics are local minima of both functionals L and E, i.e. if we consider a small perturbation η of a geodesic γ, where the support η is also small, the length of the new trajectory $\gamma + \eta$ is not less

than that of γ. We now formulate the exact problem. Consider a compact Riemannian manifold M^n. According to the results of Chapter 5, there exists an $\varepsilon > 0$ such that any pair of points P, Q

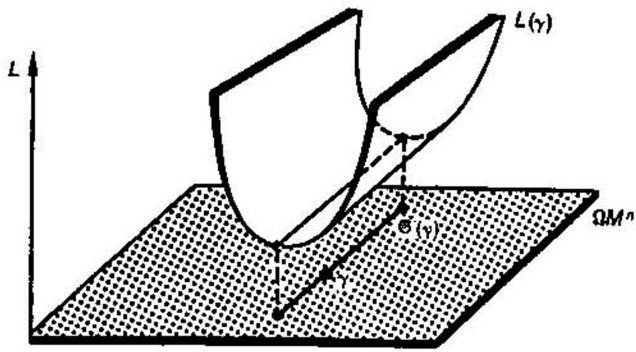

Figure 7.8

in the ball D^n_ε of radius ε can be connected by a unique geodesic entirely lying in the ball.

Theorem 3. *Let D^n_ε and ε be the ball and the number just mentioned, and let γ: $[0, 1] \to M^n$ be a geodesic of length less than ε which connects two points in D^n_ε. Suppose ω: $[0, 1] \to M^n$ is another smooth path connecting the same points (the path can also be piecewise-smooth). Then $L(\omega) \geqslant L(\gamma)$, the equality taking place only when the point sets $\gamma[0, 1]$ and $\omega[0, 1]$ coincide (i.e. when these two paths coincide at smooth curves in M^n). In this sense, the geodesic γ is the shortest path between the points P and Q.*

Proof. Let P_0 be the centre of the ball D^n_ε. From Chapter 5 we know that the smooth mapping $\exp_{P_0}$: $B^n_\varepsilon \to M^n$ defined by $\exp_{P_0}\mathbf{a} = \gamma_\mathbf{a}(1)$ is a diffeomorphism. Here $\mathbf{a} \in T_{P_0}M^n$, B^n_ε is the ball of radius ε in $T_{P_0}M^n$, and $\gamma_\mathbf{a}(t)$ is a geodesic in M^n such that $\gamma_\mathbf{a}(0) = P_0$, $\dot\gamma_\mathbf{a}(0) = \mathbf{a}$ (see Fig. 7.9). This mapping is called exponential. Since $T_{P_0}M^n$ is provided with metric-induced scalar product, $\exp_{P_0}$ preserves the lengths of vectors emerging from the point 0. We now prove that in D^n_ε geodesics emerging from P_0 are orthogonal to the hypersurfaces $S^{n-1} = \{\exp_{P_0}\mathbf{a},\ |\mathbf{a}| = \text{const}\}$. Indeed, let $t \to \mathbf{a}(t)$ be an arbitrary curve in $T_{P_0}M^n$ such that $|\mathbf{a}(t)| = 1$. It is required to verify that the corresponding curves in M^n, $t \to \exp_{P_0}(r_0\mathbf{a}(t))$, $0 < r_0 \leqslant \varepsilon$, are orthogonal to the geodesics $r \to \exp_{P_0}(r\mathbf{a}(t))$ (see Fig. 7.10). Let us consider a two-dimensional surface

$$f(r, t) = \exp_{P_0}(r\mathbf{a}(t)), \quad \begin{cases} 0 \leqslant r < \varepsilon, \\ t_0 - \lambda \leqslant t \leqslant t_0 + \lambda \end{cases}$$

parametrized with two parameters, r and t. This surface may be looked upon as a cone generated by the curve $f(r_0, t)$ (Fig. 7.11). It is required to prove that $\left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = 0$ at each point of this

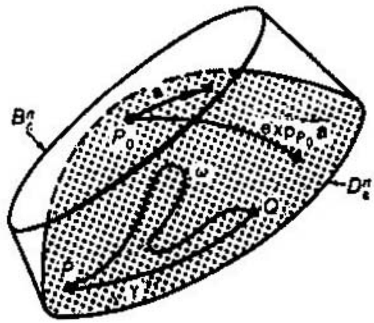

Figure 7.9

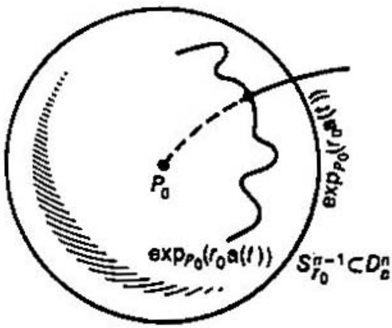

Figure 7.10

surface, where $\frac{\partial f}{\partial r}$ and $\frac{\partial f}{\partial t}$ are tangent vectors to the coordinate network on the surface. We have

$$\frac{\partial}{\partial r} \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = \nabla_r \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle,$$

where ∇_r is the covariant derivative with respect to r. Also,

$$\nabla_r \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = \left\langle \nabla_r \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle + \left\langle \frac{\partial f}{\partial r}, \nabla_r \frac{\partial f}{\partial t} \right\rangle.$$

Since the curves $r \to f(r, t)$ are geodesics and $\frac{\partial f}{\partial r}$ is the geodesic vector field, we have $\nabla_r \frac{\partial f}{\partial r} = 0$. whence $\left\langle \nabla_r \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = 0$. Further-

more, $\left\langle \frac{\partial \mathbf{f}}{\partial r}, \nabla_r \frac{\partial \mathbf{f}}{\partial t} \right\rangle = \left\langle \frac{\partial \mathbf{f}}{\partial r}, \nabla_t \frac{\partial \mathbf{f}}{\partial r} \right\rangle$, since $\nabla_t \left(\frac{\partial}{\partial r} \right) = \nabla_r \left(\frac{\partial}{\partial t} \right)$, i.e. $\nabla_{\frac{\partial}{\partial t}} \left(\frac{\partial}{\partial r} \right) = \nabla_{\frac{\partial}{\partial r}} \left(\frac{\partial}{\partial t} \right)$ because the fields $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial t}$

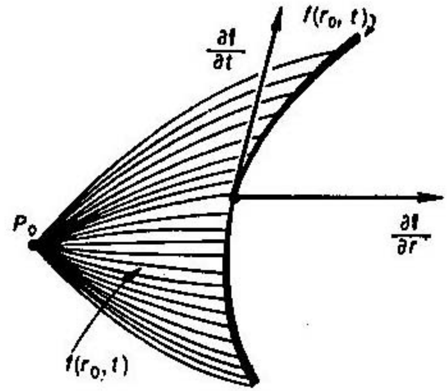

Figure 7.11

commute (see Chapter 5, Sec. 5.5). Hence,

$$\left\langle \frac{\partial \mathbf{f}}{\partial r}, \nabla_t \frac{\partial \mathbf{f}}{\partial r} \right\rangle = \frac{1}{2} \nabla_t \left\langle \frac{\partial \mathbf{f}}{\partial r}, \frac{\partial \mathbf{f}}{\partial r} \right\rangle = \frac{1}{2} \frac{\partial}{\partial r} \left\langle \frac{\partial \mathbf{f}}{\partial r}, \frac{\partial \mathbf{f}}{\partial r} \right\rangle = 0,$$

since $\left| \frac{\partial \mathbf{f}}{\partial r} \right|^2 = \left\langle \frac{\partial \mathbf{f}}{\partial r}, \frac{\partial \mathbf{f}}{\partial r} \right\rangle = \text{const}$ and $\frac{\partial \mathbf{f}}{\partial r}$ is the velocity vector of the geodesic $r \to f(r, t)$ (see (Fig. 7.11).

Thus, the function $\left\langle \frac{\partial \mathbf{f}}{\partial r}, \frac{\partial \mathbf{f}}{\partial t} \right\rangle$ does not depend on r. But for $r = 0$ we have $f(0, t) = \exp_{P_0} 0 = P_0$, i.e. $\frac{\partial \mathbf{f}(0, t)}{\partial t} = 0$, whence $\left\langle \frac{\partial \mathbf{f}}{\partial r}, \frac{\partial \mathbf{f}}{\partial t} \right\rangle = 0$ for all r, which is what was required. Let now $\omega\colon [a, b] \to D^n_\varepsilon \setminus P_0$ be an arbitrary smooth curve. Each point of $\omega(t)$ can uniquely be represented in the form $\exp_{P_0}(r(t)\,\mathbf{a}(t))$, where $0 < r(t) < \varepsilon$ and $|\mathbf{a}(t)| = 1$, $\mathbf{a}(t) \in T_{P_0}M^n$. We now prove that $\int\limits_a^b |\dot{\omega}(t)|\, dt \geqslant |r(b) - r(a)|$, the equality taking place if only the function $r(t)$ is monotonic and the function $\mathbf{a}(t)$ is constant. Once this fact is proved, we immediately obtain that a radial

geodesic is the shortest curve connecting two concentric spheres with centre at P_0 (see Fig. 7.12). Recall that $f(r, t) = \exp_{P_0}(r \cdot \mathbf{a}(t))$ and, therefore, $\omega(t) = f(r(t), t)$. We have $\frac{d\omega}{dt} = \frac{\partial f}{\partial r}\mathbf{r}'(t) + \frac{\partial \mathbf{f}}{\partial t}$. Since $\mathbf{r}'(t)$ and $\frac{\partial \mathbf{f}}{\partial t}$ are orthogonal (see above) and $\left|\frac{\partial \mathbf{f}}{\partial r}\right| = 1$, we obtain

$$\left|\frac{\partial\omega}{\partial r}\right|^2 = |\mathbf{r}'(t)|^2 + \left|\frac{\partial \mathbf{f}}{\partial t}\right|^2 \geqslant |\mathbf{r}'(t)|^2,$$

the equality taking place if only $\left|\frac{\partial \mathbf{f}}{\partial t}\right| = 0$, i.e. for $\frac{d\mathbf{a}(t)}{dt} = 0$. Thus,

$$\int_a^b \left|\frac{d\omega}{dt}\right| dt \geqslant \int_a^b |\mathbf{r}'(t)|\, dt \geqslant |r(b) - r(a)|,$$

the equality taking place if only $r(t)$ is monotonic and $\mathbf{a}(t)$ is constant. The inequality is proved.

We now turn to the proof of the theorem. Let $\omega(t)$ be an arbitrary smooth path from P_0 to $P = \exp_{P_0}(r\mathbf{a}) \in D^n_{\varepsilon_1}$, where $0 < r < \varepsilon$,

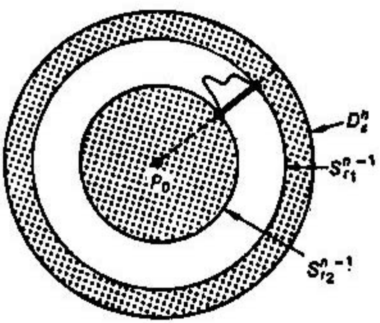

Figure 7.12

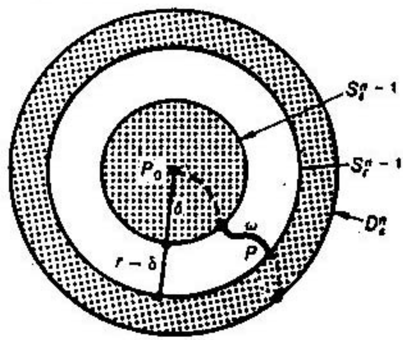

Figure 7.13

$|\mathbf{a}| = 1$. Then for any $\delta > 0$ the path $\omega(t)$ must include a smooth segment from the sphere S^{n-1}_δ of radius δ to the sphere S^{n-1}_r of radius r (see Fig. 7.13). According to what has been proved above, the length of this segment is not less than $r - \delta$. By making δ tend to zero, we find that the length of ω is not less than r. On the other hand, the segment of the geodesic from P_0 to P is of length r. The theorem is proved.

Corollary 1. *Let ω: $[0, \varepsilon] \to M^n$ be a smooth trajectory parametrized by the natural parameter and let the length of the path from $P_0 = \omega(0)$ to $\omega(\varepsilon)$ not exceed the length of any other path from P_0 to $\omega(\varepsilon)$. Then ω is a geodesic.*

Proof. Consider an arbitrary path inside the ball D^n_ε defined above, where P_0 is the beginning of the path. Then the statement follows from Theorem 3.

Definition. A geodesic γ: $[a, b] \to M^n$ is called *minimal* if it is shorter than any other smooth path connecting its extreme points $\gamma(a)$ and $\gamma(b)$.

Theorem 3 asserts that any sufficiently small segment of a geodesic is minimal. At the same time, a rather long geodesic may not be minimal. For example, we have proved above that any equator on the sphere S^2 is a geodesic. Consider the segment SNP of this equator shown in Fig. 7.14. Apparently, the geodesic SNP is not minimal because the segment $P\alpha S$ of the equator is shorter than SNP.

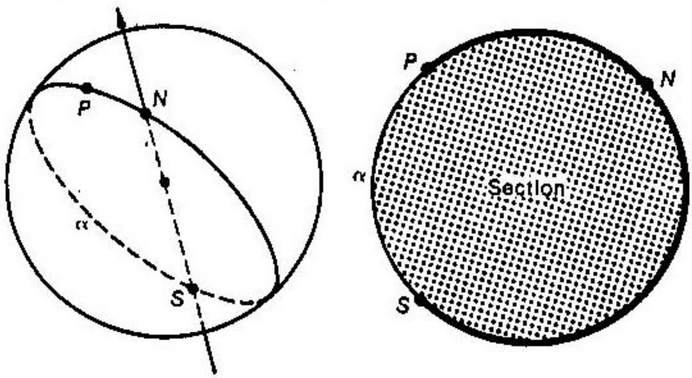

Figure 7.14

A minimal (long) geodesic connecting two points P and Q need not necessarily be unique. For instance, the north and south poles on a two-dimensional sphere can be connected by infinitely many minimal geodesics, meridians.

As another corollary of Theorem 3, we now demonstrate how to find all geodesics on an n-dimensional sphere S^n (n is arbitrary). We might, of course, proceed by analogy with the two-dimensional case by writing out all Christoffel symbols and solving the equations of geodesics. However, we shall handle the problem more elegantly.

Corollary 2. *A curve on sphere S^n endowed with the standard metric is a geodesic if and only if this curve is an equator* (*i.e. a section of S^n by a two-dimensional plane through the sphere centre*).

Proof. Let us consider an arbitrary two-dimensional plane $\mathbf{R}^2$ intersecting S^n along the equator γ. Let g be a reflection in $\mathbf{R}^n$ which preserves $\mathbf{R}^2$ fixed. Then there exists the isometry g: $S^n \to S^n$ such that the set of its fixed points exactly coincides with the equator γ. Let P and Q be sufficiently close points on γ which are connected

by a unique minimal geodesic ω (see Theorem 3). Since g is an isometry, the curve $g(\omega)$ through the points P and Q is also a geodesic which connects P and Q; hence, $\omega = g(\omega)$, i.e. $\omega = \gamma$ is a geodesic. That there exist no other geodesics is proved similarly to the two-dimensional case: only one equator passes through any point of a sphere S^n in any direction. The corollary is proved.

Using exactly the same reasoning, we can prove that any meridian of a surface of revolution is a geodesic.

We consider an arbitrary Riemannian compact manifold. Let P and Q be arbitrary points and let $\Omega(P, Q)$ be the space of all smooth curves γ connecting P and Q, i.e. $\gamma(0) = P$, $\gamma(1) = Q$. Two funcionals, E and L (see above), are valid on this space. We define the distance $\rho(P, Q)$ between the points P and Q as the infimum of the lengths of smooth curves through these points.

Proposition 1. *Let P and Q be sufficiently close points of a manifold spaced at a distance d. Then the functional of action $E: \Omega(P, Q) \to \mathbf{R}$ reaches the absolute minimum d^2 on the minimal geodesic that connects P and Q.*

Remark. Since a locally minimal geodesic is unique (see above), it is represented by an isolated point of the absolute minimum. This means that the value of the functional E on any smooth curve sufficiently close (both pointwise and in the sense of velocity field) to the minimal geodesic is strictly larger than on the minimal geodesic itself.

Proof. Let γ be a minimal geodesic, $\gamma(0) = P$, $\gamma(1) = Q$. According to Lemma 1, $E(\gamma) = L^2(\gamma) \leqslant L^2(\omega) \leqslant E(\omega)$, where ω is a smooth curve from P to Q. The equality $L^2(\omega) = L^2(\gamma)$ holds true if and only if ω is also a minimal geodesic (maybe, differently parametrized to within a scale transformation). At the same time, the equality $L^2(\omega) = E(\omega)$ is valid if only the parameter is proportional to the arc length. Hence, $E(\gamma) < E(\omega)$ in all the cases where ω is not a minimal geodesic. The proposition is proved.

7.3. MINIMAL SURFACES

In Chapter 4 we considered minimal two-dimensional surfaces M^2, i.e. surfaces whose mean curvature H is zero. Let us now analyse minimal surfaces from the point of view of extremal functions for a functional of area.

Let us consider a functional of $(n-1)$-dimensional volume defined on compact hypersurfaces which are the graphs of smooth functions $x^n = f(x^1, \ldots, x^{n-1})$ with the domain D embedded in $\mathbf{R}^{n-1}(x^1, \ldots, x^{n-1})$. Suppose the domain D has a smooth boundary ∂D and is bounded (Fig. 7.15). In Chapter 5 we have learned that the $(n-1)$-dimensional volume of a hypersurface $V^{n-1} = \{x^n = f(x^1, \ldots,$

$x^{n-1})\}$ can be written as

$$\operatorname{vol} V^{n-1} = \int_D \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}\, dx^1 \ldots dx^{n-1},$$

where $f_{x^i} = \frac{\partial f}{\partial x^i}$. Since the Lagrangian L is of the form

$$L(f_{x^1}, \ldots, f_{x^{n-1}}) = \sqrt{1 + \sum_{i=1}^{n} (f_{x^i})^2}.$$

Euler's equation becomes

$$\sum_{i=1}^{n-1} \frac{\partial}{\partial x^i} \left(\frac{\partial f}{\partial x^i} \cdot \left(1 + \sum_{i=1}^{n-1} \left(\frac{\partial f}{\partial x^i} \right)^2 \right)^{-\frac{1}{2}} \right) = 0.$$

This is a differential equation satisfied by the extremal function $x^n = f(x^1, \ldots, x^{n-1})$. Let us consider an arbitrary $(n-1)$-dimensional hypersurface $V^{n-1} \subset \mathbf{R}^n$ which is an extremal function for

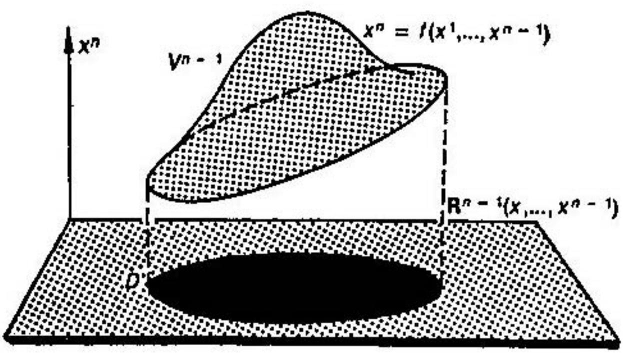

Figure 7.15

the functional of $(n-1)$-dimensional volume. Since volume is a scalar and does not depend on the choice of coordinates on the surface, we may work with particular suitable coordinates. Let V^{n-1} be an extremal surface and let $P \in V^{n-1}$ be an arbitrary point. We set $\mathbf{R}^{n-1}(x^1, \ldots, x^{n-1}) = T_P V^{n-1}$, i.e. we deal with the tangent plane with Cartesian coordinates; we also express (locally) V^{n-1} as the graph of the function $x^n = f(x^1, \ldots, x^{n-1})$.

Theorem 1. *A hypersurface $V^{n-1} \subset \mathbf{R}^n$ is extremal for the functional of $(n-1)$-dimensional volume if and only if its mean curvature H is identically zero.*

Proof. We consider the case $n = 3$, since for arbitrary n calculations are quite similar. For $M^2 \subset \mathbf{R}^3$ we have

$$\frac{\partial}{\partial x}\left(\frac{f_x}{\sqrt{1+f_x^2+f_y^2}}\right)+\frac{\partial}{\partial y}\left(\frac{f_y}{\sqrt{1+f_x^2+f_y^2}}\right)=0.$$

Differentiation yields

$$f_{xx}+f_{xx}f_x^2+f_{xx}f_y^2-f_x^2f_{xx}-f_xf_yf_{xy}+f_{yy}+f_{yy}f_x^2 \\ +f_{yy}f_y^2-f_xf_yf_{xy}-f_y^2f_{yy}=0,$$

i.e.

$$f_{xx}(1+f_y^2)-2f_xf_yf_{xy}+f_{yy}(1+f_x^2)=0.$$

By virtue of Chapter 4, this equation coincides with $H \equiv 0$.

Thus, a minimal surface can be viewed as a surface defined by an extremal radius vector.

Let us consider a surface M^2 in $\mathbf{R}^3$ referred to (u, v), i.e. $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$. Then the functional of area is written as (see Chapter 5): $S[\mathbf{r}] = \int\limits_{D(u,v)} \sqrt{EG-F^2}\,du\,dv$. Let (u, v) be conformal parameters, i.e. the parameters in which the metric on M^2 is of the form

$$E=G=\langle\mathbf{r}_u, \mathbf{r}_u\rangle=\langle\mathbf{r}_v, \mathbf{r}_v\rangle, \quad F=0.$$

Then

$$S[\mathbf{r}]=\int\limits_D (\langle\mathbf{r}_u, \mathbf{r}_u\rangle\langle\mathbf{r}_v, \mathbf{r}_v\rangle)^{1/2}\,du\,dv.$$

Euler's equations are written as

$$\frac{\partial}{\partial u}(2x_u)+\frac{\partial}{\partial v}(2x_v)=0, \quad \frac{\partial}{\partial u}(2y_u)+\frac{\partial}{\partial v}(2y_v)=0, \\ \frac{\partial}{\partial u}(2z_u)+\frac{\partial}{\partial v}(2z_v)=0,$$

i.e. $\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right)\mathbf{r}(u, v)=0$. Thus, $\mathbf{r}$ is a harmonic radius vector (relative to the coordinates (u, v)). **Hint:**

$$L(\mathbf{r}_u, \mathbf{r}_v)=\sqrt{(x_u^2+y_u^2+z_u^2)(x_v^2+y_v^2+z_v^2)}$$

and all partial derivatives are considered as independent in the variation procedure.

As in the one-dimensional case, we consider the infinite-dimensional space F of smooth mappings $D^2(u, v)$ in $\mathbf{R}^3$. On F, there is defined a non-linear functional of area $S[\mathbf{r}]$, its extremal "points" (i.e. radius vectors $\mathbf{r}(u, v)$) being described by the following theorem.

Theorem 2. *The vectors* $\mathbf{r}(u, v)$ *are extremal vectors for* $S[\mathbf{r}]$ *if and only if their mean curvature* H *is zero.*

The statement follows from Theorem 1. Let us consider one more functional, the Dirichlet functional $D[\mathbf{r}] = \frac{1}{2} \int\limits_{D(u,\,v)} (E + G)\, du\, dv$, and compare extremal "points" for the functionals D and S.

Theorem 3. *The vectors* $\mathbf{r}(u, v)$ *are extremal vectors for the Dirichlet functional, if and only if they are harmonic with respect to* (u, v), *i.e.* $\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right) \mathbf{r}(u, v) = 0$.

Proof. Euler's equations are of the form $\frac{\partial}{\partial u}\left(\frac{\partial L}{\partial \mathbf{r}_u}\right)+\frac{\partial}{\partial v}\left(\frac{\partial L}{\partial \mathbf{r}_v}\right)=0$, where $L = E + G = \langle \mathbf{r}_u, \mathbf{r}_u\rangle + \langle \mathbf{r}_v, \mathbf{r}_v\rangle$, i.e. $L = x_u^2 + x_v^2 + y_u^2 + y_v^2 + z_u^2 + z_v^2$, i.e. $\Delta\mathbf{r} = 0$, where $\Delta = \frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}$. The theorem is proved.

The coordinates (u, v) may be harmonic for a harmonic radius vector, but they may not be conformal for the induced metric on the surface swept out by this vector. We now consider a special harmonic surface on which harmonic coordinates are, at the same time, conformal, i.e. $E = G$, $F = 0$. Such a surface is, therefore, minimal. Thus, certain extremal "points" of the Dirichlet functional generate extremal "points" of the functional of area. In other words, if the initial parameters of a harmonic vector are subjected to an arbitrary regular transformation, each "harmonic point" referred to conformal coordinates gives rise to a family of minimal radius vectors. This situation resembles the interaction between extremal "points" of functionals of length and action in the one-dimensional variation problem, i.e. on the space of smooth trajectories. Not every harmonic vector generates a minimal surface. The functionals D and S satisfy the relation $D[\mathbf{r}] \geqslant S[\mathbf{r}]$ for any radius vector, the equality taking place if and only if $E = G$, $F = 0$, i.e. if the coordinate network is orthogonal and the coordinates (u, v) are conformal. The proof follows from the obvious inequality $\frac{E+G}{2} \geqslant \sqrt{EG - F^2}$ which becomes equality if and only if $F = 0$, $E = G$.

Thus far, we have proved that a minimal surface is an extremal of the functional of area, but we have not justified the choice of the term "minimal surfaces". We did so for the one-dimensional case: we proved that a geodesic is a locally minimal trajectory. A similar statement holds true for an extremal of the functional of area: a minimal surface has the property that for an arbitrary perturbation $\boldsymbol{\eta}$ of the minimal radius vector $\mathbf{r}$ with small support (i.e. perturbation is non-zero only in a small domain) the area of the "perturbed surface" $\mathbf{r} + \boldsymbol{\eta}$ is not less than the area of the initial surface. The proof of this assertion requires an additional analysis, since the vanishing of the first variation by no means ensures that the extremal "point" is

locally minimal in the space of all radius vectors. Even for an ordinary function of one argument, the vanishing of the first derivative does not show with certainty whether the function has a minimum, a maximum, or an inflection point. For example, in the case of geodesics we had to analyse thoroughly the local properties of the functionals of length and action. And this is a "second-order" analysis, just as in the case of an ordinary function of one argument for which we usually analyse the second-order differential to study the local behaviour of the function near a critical point.

While investigating local minimality of a geodesic, we considered, in fact, the so-called "second variation" of the functional of action

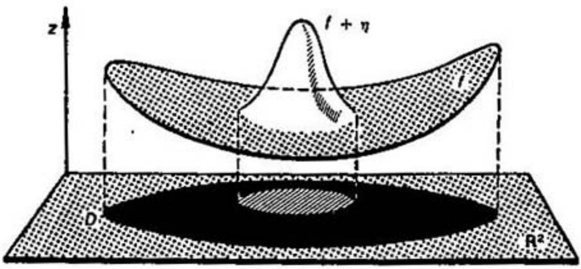

Figure 7.16

(or length), an analogue of second-order differential. Hence, in the case of extremals of the functional of area one should also study the "second derivative" of the functional. We shall not go into details of the second variation theory, but instead we shall use a simpler approach (though this approach is more special).

Let us consider a minimal surface $M^2 \subset \mathbf{R}^3$ (the reasoning is similar for an arbitrary n), $P \in M^2$, and define M^2 in a small neighbourhood of P as the graph of a smooth function $z = f(x, y)$, where the Cartesian coordinates (x, y) vary in the tangent plane to M^2 at P. We now prove that any sufficiently small perturbation (with small support) of M^2 does not reduce the area. Let in $\mathbf{R}^3$ there be given the graph $z = f(x, y)$ defined over a domain D in the plane $\mathbf{R}^2(x, y)$; the surface is minimal, i.e. $H = 0$ and we consider the perturbation of the graph that vanishes at the boundary of D. It is required to prove that $S[f] \leqslant S[f + \eta]$ (see Fig. 7.16).

Let us consider the space $F(f)$ of all smooth functions $f + \eta$ defined on D and such that $\eta|_{\partial D} = 0$. This linear space depends on the choice of the function f. The space $F(f)$ is obtained from the linear space C of smooth functions η vanishing at the boundary, i.e. "per-

turbations" of the function f due to shifting C by the function η (see Fig. 7.17). Consider the restriction of the functional S to $F(f)$. On this space, the functional S associates with each function $f + \eta$ the areas of its graph. Let us prove that this functional is "convex

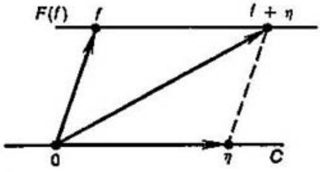

Figure 7.17

downward", i.e. $S[\alpha r + \beta g] \leqslant \alpha S[r] + \beta S[g]$, where $\alpha + \beta = 1$; $r, g \in F(f)$, i.e. $r = f + \eta_1$; $g = f + \eta_2$, where $\eta_i|_{\partial D} = 0$, $i = 1, 2$. Note that $\alpha r + \beta g \in F(f)$ because

$$(\alpha r + \beta g)|_{\partial D} = \alpha r|_{\partial D} + (1 - \alpha) g|_{\partial D}$$
$$= \alpha f|_{\partial D} + (1 - \alpha) f|_{\partial D} = f|_{\partial D},$$

i.e. $\alpha r + \beta g = f + \eta_3$, where $\eta_3|_{\partial D} = 0$. The definition of a convex-downward functional copies the analogous definition for an ordinary

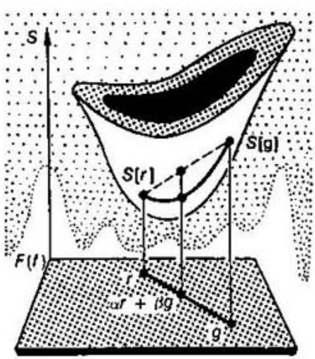

Figure 7.18

function. The graph of a convex-downward functional is shown conditionally in Fig. 7.18. Thus, it suffices to verify that

$$\sqrt{1 + (\alpha r_x + \beta g_x)^2 + (\alpha r_y + \beta g_y)^2}$$
$$\leqslant \alpha \sqrt{1 + r_x^2 + r_y^2} + \beta \sqrt{1 + g_x^2 + g_y^2}.$$

Squaring this inequality, we obtain

$$1+2\alpha\beta(r_x g_x + r_y g_y) \leqslant \alpha^2+\beta^2+2\alpha\beta\sqrt{(1+r_x^2+r_y^2)(1+g_x^2+g_y^2)},$$

and since $1-\alpha^2-\beta^2=2\alpha\beta$,

$$1+r_x g_x + r_y g_y \leqslant \sqrt{(1+r_x^2+r_y^2)(1+g_x^2+g_y^2)},$$

or $\langle \mathbf{a}, \mathbf{b}\rangle \leqslant |\mathbf{a}|\cdot|\mathbf{b}|$, where

$$\mathbf{a}=(r_x, r_y, -1)=\operatorname{grad}(r(x, y)-z),$$
$$\mathbf{b}=(g_x, g_y, -1)=\operatorname{grad}(g(x, y)-z).$$

The inequality $\langle \mathbf{a}, \mathbf{b}\rangle \leqslant |\mathbf{a}|\,|\mathbf{b}|$ is obvious. Thus, we have proved that S is convex downward on $F(f)$, whence it immediately follows that any extremal "point" in the space $F(f)$ is a minimum for the functional S. In particular, for all points r in some neighbourhood of an extremal point $g \in F(f)$ the inequality $S[r] \geqslant S[g]$ is satisfied. Since the point $f \in F(f)$ is extremal, we have $S[f] \leqslant S[f+\eta]$, $\eta|_{\partial D}=0$. Thus, we have proved the following statement.

Theorem 4. *Let $M^2 \subset \mathbf{R}^3$ be an arbitrary minimal surface, then this surface is locally minimal, i.e. any smooth, sufficiently small perturbation with small support does not diminish the area of the surface.*

The theorem on the local minimality of extremal solutions for the functional of an $(n-1)$-dimensional volume in $\mathbf{R}^n$ is proved similarly.

7.4. CALCULUS OF VARIATIONS AND SYMPLECTIC GEOMETRY

Let T_*M^n be a tangent bundle of a manifold M^n, i.e. a $2n$-dimensional smooth manifold whose points are represented by pairs $\{(x, \mathbf{a})\}$, where $x \in M^n$, $\mathbf{a} \in T_*M^n$. Also, let $P, Q \in M^n$ and let $\Omega(P, Q)$ be the space of all smooth curves connecting P and Q. Let local coordinates $x^1, \ldots, x^n$ be valid on M^n in which a trajectory is written in the form $\gamma(t) = (x^1(t), \ldots, x^n(t))$, $x(0) = P$, $x(1) = Q$; $\dot{x}(t)$ is the velocity vector. Consider the functional $I[\gamma] = \int\limits_0^1 L(x, \dot{x})\,dt$, where L is a smooth function of two groups of variables, $(x, \dot{x}) = \mathbf{a}$, i.e. $L(x, \mathbf{a})$ is a function on T_*M^n.

Definition. The *momentum* $p=(p_i)$, $1 \leqslant i \leqslant n$, is the covector with the components $p_i = \dfrac{\partial L}{\partial \dot{x}^i}$ (in a given coordinate system). The *energy* E is the function $E(x, \dot{x}) = \dot{x}^i p_i - L(x, \dot{x})$.

The energy can be considered as a function on T_*M^n, i.e. $E(x, \mathbf{a}) = a^i p_i - L(x, \mathbf{a})$. An important example: $L(x, \dot{x}) = g_{ij}\dot{x}^i\dot{x}^j$, in this

case $p_i = \frac{\partial L}{\partial \dot{x}^i} = 2g_{ij}\dot{x}^j$, i.e. p is the covector dual to the velocity vector $\mathbf{a} = \dot{x}$ relative to the metric g_{ij}. The energy E is of the form

$$E(x, \dot{x}) = \dot{x}^i 2g_{ij}\dot{x}^j - g_{ij}\dot{x}^i\dot{x}^j = g_{ij}\dot{x}^i\dot{x}^j,$$

i.e. it is "kinetic energy".

Let $\mathfrak{G}$ be a Lie group which acts smoothly on $M^n(x^1, \ldots, x^n)$ (i.e. each element $g \in \mathfrak{G}$ is represented by a diffeomorphism of M^n). We shall focus on the following example: $\mathfrak{G} = \mathbf{R}^1$, i.e. the real straight line (the group with respect to addition). This action (for $\mathfrak{G} = \mathbf{R}^1$) generates on M^n a vector field $\mathbf{X}(x_0) = \frac{d}{dt} g_t(x_0)|_{t=0}$, where the orbit $\mathfrak{G}(x_0) = g_t(x_0)$; $g(x_0) = x_0$. The action of $\mathfrak{G}$ on M^n induces the action of $\mathfrak{G}$ on T_*M^n according to the mapping g_*: $(x, \mathbf{a}) \to (g(x), dg_x(\mathbf{a}))$.

Definition. The Lagrangian $L(x, \mathbf{a})$ is said to be *preserved under the action* of $\mathfrak{G}$ (or *is invariant with respect to* $\mathfrak{G}$) if the action of $\mathfrak{G}$ on T_*M^n transforms the function $L(x, \mathbf{a})$ into itself, i.e.

$$L(g(x), dg(\mathbf{a})) \equiv L(x, \mathbf{a}), \quad (x, \mathbf{a}) \in T_*M^n.$$

We now determine analytic conditions for L to be invariant. The condition of the preservation of L is $\frac{dL}{dt} = 0$, where t is time along the trajectory $g_t(x)$. The differentiation of the composite function yields

$$\frac{dL}{dt} = \frac{\partial L}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial L}{\partial a^i}\frac{da^i}{dt} = 0, \quad \frac{dx^i}{dt} = X^i(x)$$

(see above). Let us calculate $\frac{da^i}{dt}$. Consider $\mathbf{a}$ and a tangent trajectory $\varphi(\tau) = (x^i(\tau))$ such that $a^i = \frac{dx^i}{d\tau}\Big|_{\tau=0}$. A small shift by Δt along $g_t(x)$ transforms the functions x^i into $x^i + X^i\,\Delta t$ (small quantities of order higher than one are neglected). It follows that

$$\frac{dx^i}{d\tau} \to \frac{dx^i}{d\tau} + \frac{\partial X^i}{\partial x^k}\,\Delta t\,\frac{dx^k}{dt}, \quad \text{i.e. } a^i \to a^i + \frac{\partial X^i}{\partial x^k}\,a^k\Delta t,$$

$$\text{i.e. } \frac{da^i}{dt} = \frac{\partial X^i}{\partial x^k}\,a^k,$$

and therefore $g_{\Delta t}(x^i) = \{x^i + X^i\Delta t + \ldots\}$,

$$\frac{\partial}{\partial x^k}(g_{\Delta t}(x^i)) = \frac{\partial}{\partial x^k}(x^i + X^i\Delta t + \ldots) \cong \left(\delta^i_x + \frac{\partial X^i}{\partial x^k}\,\Delta t\right) = J^i_k,$$

$J=(J_k^j)$ is the Jacobi matrix, whence

$$a^j \to J_k^j a^k = a^j + \frac{\partial X^j}{\partial x^k} a^k \Delta t.$$

Thus,

$$\frac{dL}{dt} = \frac{\partial L}{\partial x^i} X^i + \frac{\partial L}{\partial a^l} \frac{\partial X^l}{\partial x^k} a^k \equiv 0,$$

where $(x, \mathbf{a})$ are independent arguments. We have proved the following statement.

Statement 1. *The analytic condition for the Lagrangian $L(x, \mathbf{a})$ to be preserved under the action of the group $\mathfrak{G} = \mathbf{R}^1$ is*

$$\frac{\partial L}{\partial x^i} X^i + \frac{\partial L}{\partial a^l} \cdot \frac{\partial X^l}{\partial x^k} a^k = 0,$$

where $(x, \mathbf{a})$ are independent arguments and $\mathbf{X}$ is the velocity vector field due to the action of $\mathfrak{G} = \mathbf{R}^1$.

We now derive the so-called law of conservation of momentum projection along the extremals of the functional $J[\gamma]$. Let us consider $\int_0^1 L(x, \dot{x})\, dt = J[\gamma]$ and let γ_0 be an extremal of $J[\gamma]$, i.e. a solution of the system of Euler's equations $\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial \tau}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = 0$, where τ denotes time along the extremal. Let a dot stand for differentiation with respect to the parameter τ (along the extremal), then Euler's equations take the form $\frac{\partial L}{\partial x^i} = \dot{p}_i$, where p is the momentum. Consider the contraction $f(\tau) = (p, \mathbf{X}) = X^i p_i$, where $f(\tau)$ is a smooth function along $\gamma_0(\tau)$. Here $f(\tau)$ may be assumed to be the value of the covector p on the vector $\mathbf{X}$, i.e. $f(\tau) = p(\mathbf{X})$.

Theorem 1. *The identity $(f(\tau))^{\cdot} \equiv 0$ is valid, i.e. $(p, \mathbf{X})^{\cdot} \equiv 0$.*

Proof. From Statement 1 we have

$$(X^i p_i)^{\cdot} = \dot{X}^i p_i + X^i \dot{p}_i = X^i \frac{\partial L}{\partial x^i} + \frac{\partial X^l}{\partial x^k} \frac{\partial L}{\partial a^l} \frac{dx^k}{d\tau}$$

$$= \frac{\partial L}{\partial x^i} X^i + \frac{\partial X^l}{\partial x^k} \frac{\partial L}{\partial a^k} a^k = 0,$$

which is what was required. The theorem is proved.

Let us consider our model example: $L(x, \dot{x}) = g_{ij}\dot{x}^i\dot{x}^j$; then $p_i = 2g_{ij}\dot{x}^j$, i.e. $X^i p_i = 2g_{ij}X^i\dot{x}^j = 2\langle \mathbf{X}, \dot{\gamma}_0\rangle = \text{const}$ along $\gamma_0(\tau)$.

Extremals of this Lagrangian are geodesics (see Sec. 7.2); in particular, $|\dot{\gamma}_0| = \text{const}$ along $\gamma_0(\tau)$, whence $|\dot{\mathbf{X}}| \cos\alpha(\tau) = \text{const}$, where $\alpha(\tau)$ is the angle between the vectors $\mathbf{X}$ and $\dot{\gamma}_0$ (measured in the metric g_{ij}).

We now use this result to find geodesics on a surface of revolution M^2 in $\mathbf{R}^3$. Let $r(z)$ be a smooth function which is the generator of the surface of revolution, i.e. the curve $y = r(z)$ rotates around the axis Oz (Fig. 7.19). If $\omega = \text{const}$ is the speed of rotation, there arises

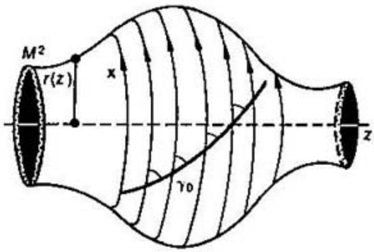

Figure 7.19

on M^2 the action $r\omega t$ of the group $\mathfrak{G} = \mathbf{R}^1$, and the field $\mathbf{X}$ of the velocities of this action has the modulus $|\mathbf{X}| = r\omega$. In view of what has been proved above, the identity $r(z)\cos\alpha(z) = c = \text{const}$ holds true for any geodesic $\gamma_0(\tau)$ on M^2. Thus, we have proved the following statement.

Statement 2. *Given a surface of revolution $M^2 \subset \mathbf{R}^3$ with the generator $r(z)$. Then the identity $r(z)\cos\alpha(z) = \text{const}$ is satisfied along any geodesic $\gamma_0(\tau)$ on M^2.*

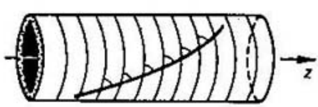

Figure 7.20

Let us consider several examples. Let $r(z) = \text{const}$, then the surface of revolution is a cylinder, and geodesics on the cylinder are the images of straight lines after the Euclidean plane is rolled into a cylinder (Fig. 7.20).

Let $r(z)$ define S^2 (Fig. 7.21). In this case geodesics are equators (the angle is variable).

The equality $r(z)\cos\alpha(z) = \text{const}$ is the necessary (but not sufficient) condition for a trajectory to be a geodesic (give an example!). Indeed, if $\gamma(\tau)$ is an orbit of the action of $\mathfrak{G}$, then $\alpha(\tau) = 0$, i.e.

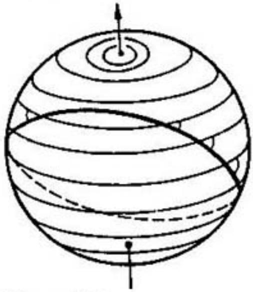

Figure 7.21

$(p, X) = |X|\,|p| = \text{const}$ (on the surface of revolution), but this orbit is not always a geodesic. Let us consider, for example, a right circular cone and its rotations about the axis (Fig. 7.22). As we

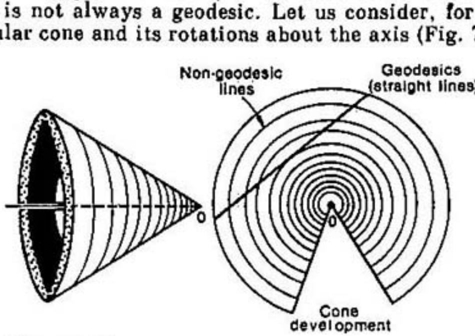

Figure 7.22

already know, all geodesics on a cone are the images of straight lines after a piece of Euclidean plane is rolled up into a cone; circles are not, apparently, geodesics on a cone.

Remark. Since $\frac{dL}{dt} \equiv 0$ along the orbits of the operation of $\mathfrak{G} = \mathbf{R}^1$ (for an invariant Lagrangian), we can locally choose a coordinate x^n along the orbits $\mathfrak{G}(x)$. We have then $(p, X)^{\cdot} \equiv 0$, which is

equivalent to $p_n = 0$, where p_n is the momentum projected onto the coordinate x^n. In other words, in this case the momentum p does not depend on x^n, and the function L does not depend on this coordinate either.

We now prove the so-called law of conservation of energy along an extremal. Let $L(x, \mathbf{a})$ denote the Lagrangian (not necessarily invariant under the action of a group) and let $\gamma_0(\tau)$ be an extremal of the functional $J[\gamma] = \int L(x, \dot{x})\,d\tau$. Consider the energy $E = \dot{x}^i p_i - L$, $p_i = \frac{\partial L}{\partial \dot{x}^i}$, and suppose $L = g_{ij}\dot{x}^i\dot{x}^j - U(x)$, where the function $U(x)$ is called a potential.

Theorem 2. *The identity* $\frac{dE}{d\tau} \equiv 0$ *holds true along extremals* $\gamma_0(\tau)$ *of the functional* $J[\gamma]$.

Proof. We have

$$\begin{aligned}\frac{dE}{d\tau} &= (p_i\dot{x}^i - L)^{\cdot}_{\tau} = (p_i a^i - L)^{\cdot}\\ &= \dot{p}_i a^i + p_i \dot{a}^i - \frac{\partial L}{\partial x^i}\dot{x}^i - \frac{\partial L}{\partial a^i}\dot{a}^i\\ &= \frac{\partial L}{\partial x^i} a^i + \frac{\partial L}{\partial a^i}\dot{a}^i - \frac{\partial L}{\partial x^i} a^i - \frac{\partial L}{\partial a^i}\dot{a}^i \equiv 0,\end{aligned}$$

since $\dot{p}_i = \frac{\partial L}{\partial x^i}$, according to Euler's equations, and $p_i = \frac{\partial L}{\partial a^i}$ by definition. The theorem is proved.

We now turn to studying simple concepts of Hamiltonian mechanics on a manifold.

Let T_*M^n be a tangent bundle of a manifold provided with coordinates $(x, \mathbf{a})$ and let L be the Lagrangian on T_*M^n.

Definition. The Lagrangian $L(x, \mathbf{a})$ is called *non-singular* if the equation

$$\mathbf{p} = \frac{\partial L(x, \mathbf{a})}{\partial \mathbf{a}} \left(\text{i.e. } p_i = \frac{\partial L(x, \mathbf{a})}{\partial a^i},\ 1 \leqslant i \leqslant n\right)$$

has a unique solution $\mathbf{a} = \mathbf{a}(x, p)$ for any $x \in M^n$.

Besides T_*M^n, we also consider a cotangent bundle T^*M^n, i.e. a $2n$-dimensional smooth manifold whose points are represented by pairs (x, p), where $p \in T^*_xM^n$, i.e. p is a covector. Let us consider the energy

$$E(x, \mathbf{a}) = p_i a^i - L(x, \mathbf{a}) = \frac{\partial L(x, \mathbf{a})}{\partial a^i} a^i - L$$

with $p = p(x, \mathbf{a})$. By expressing $\mathbf{a} = \mathbf{a}(x, p)$ from $p = p(x, \mathbf{a})$ and substituting into E, we obtain the function $E(x, \mathbf{a}(x, p)) =$

$H(x, p)$ called the *Hamiltonian*. It is of the form $H = p_i a^i(x, p) - L(x, \mathbf{a}(x, p))$, i.e. $L(x, \mathbf{a}(x, p)) = p_i a^i(x, p) - H(x, p)$. Let $\gamma(t) \subset M^n$ be a smooth trajectory, then $(x, \dot{x}) \in T_* M^n$, where $\dot{x} = \dot{x}(t)$, $x = x(t)$, i.e. $\gamma(t)$ generates a trajectory $\Gamma(t) = (x(t), \dot{x}(t))$ on $T_* M^n$. Let us now consider the Lagrangian $L(x, \mathbf{a})$ on all trajectories of $T_* M^n$, not only on trajectories $\Gamma(t) = (x(t), \dot{x}(t))$ (not every curve on $T_* M^n$ is of such a form).

We note that $T_* M^n$ and $T^* M^n$ are diffeomorphic. Indeed, let the metric g_{ij} be defined on M^n, then there exists the invariant iden-

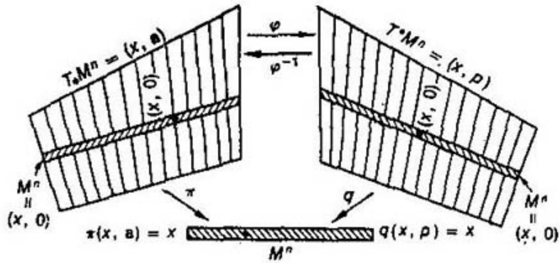

Figure 7.23

tification $\varphi\colon T_* M^n \to T^* M^n$; $\varphi(\{x^i\}, \{a^\alpha\}) = (\{x^i\}, \{g_{\beta\alpha} a^\alpha\}) \in T^* M^n$. Besides the identification φ, there exists another identification which arises in defining a non-singular Lagrangian $L(x, \mathbf{a})$. Indeed, the equation $p = \frac{\partial L(x, \mathbf{a})}{\partial \mathbf{a}}$ has a unique solution $\mathbf{a} = \mathbf{a}(x, p)$ (for a fixed x), so that with each pair $(x, \mathbf{a})$ we can associate the pair (x, p), where $p = \frac{\partial L(x, \mathbf{a})}{\partial \mathbf{a}}$. Let us consider this identification of $T_* M^n$ and $T^* M^n$, which is a diffeomorphism. In a particular case of $L(x, \mathbf{a}) = g_{ij} a^i a^j$ we have $p_i = 2 g_{ik} a^k$, i.e. the identification by this Lagrangian coincides with "Riemannian identification" (see above) (Fig. 7.23).

We now turn to the variational problem of finding extremals of the functional $I[\alpha] = \int_R^S L(x(t), \mathbf{a}(t))\, dt$, where $L(x, \mathbf{a})$ is assumed to be defined on all smooth trajectories $\alpha(t)$ in $T_* M^n$, and R and S are two fixed points on $T_* M^n$, i.e. $R(x_0, \mathbf{a}_0)$ and $S(y_0, \mathbf{b}_0)$ (Fig. 7.24). Let us analyse those variations which do not shift the points R and S. Euler's equations on $T_* M^n$ are of the form $\frac{\partial L}{\partial x^i} = \frac{d}{dt}\left(\frac{\partial L}{\partial a^i}\right)$. To

write them in terms of T^*M^n, we make the substitution $(x, p) \to (x, \mathbf{a})$ via the diffeomorphism constructed above by the Lagrangian L. We have

$$L(x, \mathbf{a}(x, p)) = p_i a^i(x, p) - H(x, p), \quad p_i = \frac{\partial}{\partial a^i} L(x, \mathbf{a}(x, p)).$$

It should be noted that $\frac{\partial}{\partial p_i} L(x, \mathbf{a}(x, p)) = 0$ since $L(x, \mathbf{a})$ does not contain the momentum p explicitly, and in the case of Euler's

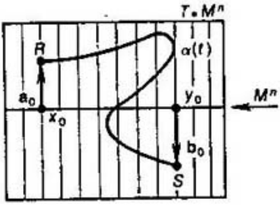

Figure 7.24

equations variations are calculated only with respect to the functions that enter explicitly into the Lagrangian. Thus,

$$0 = \frac{\partial}{\partial p_i} L(x, \mathbf{a}(x, p)) = a^i(x, p) - \frac{\partial}{\partial p_i} H(x, p),$$

where

$$L(x, \mathbf{a}(x, p)) = \left(\frac{\partial}{\partial a^i} L(x, \mathbf{a}(x, p))\right) a^i(x, p) - H(x, p),$$

i.e. $a^i(x, p) = \frac{\partial H(x, p)}{\partial p_i}$. Hence

$$\int_A^B L(x, \mathbf{a}(x, p))\, dt = \int_A^B [p_i a^i(x, p) - H(x, p)]\, dt$$

$$= \int_A^B \left[p_i \frac{\partial H(x, p)}{\partial p_i} - H(x, p)\right] dt = J[\omega(t)],$$

where $\omega(t) \in T^*M^n$, $\omega(t) = (x(t), p(t))$, $\omega(0) = A$, $\omega(1) = B$ (see Fig. 7.25). Let us now write Euler's equations for the functional $J[\omega(t)]$ on T^*M^n. Suppose $\omega_0(t) = (x(t), p(t))$ is an extremal of the functional $J[\omega(t)]$. It was shown above that $a^\alpha = \frac{\partial H(x, p)}{\partial p_\alpha}$,

but since $a^\alpha = \dot{x}^\alpha$ (along an extremal), we have $\dot{x}^\alpha = \frac{\partial H(x, p)}{\partial p_\alpha}$ (along the extremal). Furthermore,

$$J[\omega] = \int_A^B (\dot{p}_i a^i - H)\, dt = \int_A^B L(x, \mathbf{a}(x, p))\, dt.$$

Euler's equations take the form $\frac{\partial L}{\partial x_i} = \frac{d}{dt}\left(\frac{\partial L}{\partial a^i}\right)$, whence $\dot{p}_i = \frac{\partial L}{\partial x^i}$ because $\frac{\partial L}{\partial a^i} = p_i$. Also, $\frac{\partial L}{\partial x^i} = \frac{\partial}{\partial x^i}(a^i p_i - H(x, p)) = -\frac{\partial H(x, p)}{\partial x^i}$

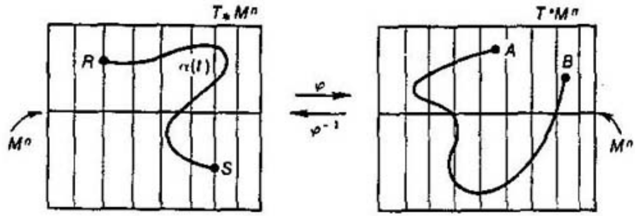

Figure 7.25

because $\frac{\partial}{\partial x^i}(a^i p_i) = 0$. While proving the relation $\frac{\partial}{\partial x^i}(a^i p_i) = 0$, we use the fact that the coordinates x and $\mathbf{a}$ in the pair $(x, \mathbf{a})$ are independent and, therefore, $\frac{\partial \mathbf{a}}{\partial x} = 0$. Moreover, we calculate the variation with respect to x^i if only this variable enters explicitly into the sum $a^i p_i$. But since it does not we have $\frac{\partial}{\partial x^i}(a^i p_i) = 0$, which is what was required. Hence $\dot{p}_i = -\frac{\partial H(x, p)}{\partial x^i}$. Thus, we have proved the following important statement.

Theorem 3. *In the coordinates* (x, p) *on the cotangent bundle* T^*M^n *Euler's equations for the functional* $J[\omega] = \int_A^B L(x, \mathbf{a}(x, p))\, dt$ *can be written in the so-called Hamiltonian form:* $\dot{x}^i = \frac{\partial H(x, p)}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H(x, p)}{\partial x^i}$ *(Hamilton's equations).*

Let us make some comments on the implicit form of the variable x^i in the sum $a^i p_i$. For example, if $L = g_{ij} a^i a^j$, then a^i is expressed in terms of the momentum p alone (x does not appear explicitly):

$p_i = \frac{\partial L}{\partial a^i} = 2g_{ij}a^j$, whence $a^\alpha = \frac{1}{2} g^{\alpha k} p_k$. Of course, the functions $g^{\alpha k} = g^{\alpha k}(x)$ do depend on x, but this variable does not appear explicitly in a^α, and therefore there is no variation with respect to x.

We now turn to studying Hamilton's equations on T^*M^n. They define on T^*M^n a smooth vector field referred to the coordinates (x^i, p_j), $1 \leqslant i, j \leqslant n$. This field is a flow of special form; not every vector field on T^*M^n can be represented as $\dot{x}^i = \frac{\partial H(x, p)}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H(x, p)}{\partial x^i}$. This field is closely related to the field grad H which is a covector field on M^n. Indeed, consider on M^n an exterior differential form of degree two $\omega^{(2)} = dp_i \wedge dx^i$ referred to the aforementioned local coordinates. At each point (x, p) this form can be represented as the skew-symmetric matrix

$$\left(\begin{array}{cc|c|cc} 0 & 1 & & & \\ -1 & 0 & & 0 & \\ \hline & & \ddots & & \\ \hline & 0 & & 0 & 1 \\ & & & -1 & 0 \end{array}\right).$$

and, therefore, defines a skew-symmetric, non-singular scalar product both in $T_{(x, p)}(T^*M^n)$ and $T^*_{(x, p)}(T^*M^n)$. Hence, this scalar product defines identification of the tangent $T_{(x, p)}(T^*M^n)$ and cotangent $T^*_{(x, p)}(T^*M^n)$ spaces at each point. Let us consider the flow grad $H(x, p)$ with the coordinates $\left(\frac{\partial H(x, p)}{\partial x^i}, \frac{\partial H(x, p)}{\partial p_i}\right)$, then the dual flow relative to this scalar product is of the form $\left(\frac{\partial H(x, p)}{\partial p_i}, -\frac{\partial H(x, p)}{\partial x}\right)$, i.e. coincides with the Hamiltonian flow derived in Theorem 3. Hence, the Hamiltonian field can be defined as a flow dual, relative to an exterior 2-form, to the gradient of the Hamiltonian $H(x, p)$ given on T^*M^n. Let us now assume M^n to be a Riemannian manifold, so that T^*M^n and T_*M^n may be considered as canonically identified, in particular, we shall not distinguish between vectors and covectors and use only upper indices in coordinate expressions.

The material presented below does not presume that the reader is familiar with the previous analysis of the functional $J[\omega]$; it merely stresses the importance of Hamiltonian flows in Riemannian geometry.

Let us consider an even-dimensional Riemannian manifold M^{2n}; in the preceding example M^{2n} was represented by $T^*M^n \cong T_*M^n$.

Let $\omega^{(2)} = \omega_{ij}\, dx^i \wedge dx^j$ be a non-singular exterior 2-form on M^{2n} which defines a skew-symmetric scalar product. If this product is given on M^{2n}, for any smooth function $f(x)$ on M^{2n} we can define a "skew-symmetric gradient", sgrad $f(x)$.

Definition. The *skew-symmetric gradient* sgrad f of a smooth function f (relative to the scalar product ω) is a vector field uniquely defined by $\omega(\mathbf{Y}, \operatorname{sgrad} f) = \mathbf{Y}(f)$, where $\mathbf{Y}$ runs the set of all smooth vector fields on M^{2n} and $\mathbf{Y}(f)$ is the value of the operator $\mathbf{Y}$ on the function f.

Although the definition of sgrad f is similar, in its form, to the definition of grad f (for a symmetric scalar product), their properties differ. Note that the uniqueness of the definition of sgrad f follows from non-singularity of ω.

We recall an important definition. Let $\omega^{(k)} = \omega_{i_1, \ldots, i_k}\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ be a differential form of degree k: the form $\omega^{(k+1)} = d\omega^{(k)}$ defined by $d\omega^{(k)} = \sum\limits_{i_0} \frac{\partial \omega_{i_1 \ldots i_k}}{\partial x^{i_0}}\, dx^{i_0} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ is called an exterior differential of $\omega^{(k)}$. For the 1-form $\omega^{(1)} = \omega_i\, dx^i$ the operation d can be written as

$$d\omega^{(1)} = \frac{\partial \omega_i}{\partial x^{i_0}}\, dx^{i_0} \wedge dx^i = \left(\frac{\partial \omega_i}{\partial x^j} - \frac{\partial \omega_j}{\partial x^i} \right) dx^i \wedge dx^j, \quad (i < j).$$

Definition. A smooth manifold M^{2n} is called *symplectic* if the exterior differential 2-form $\omega^{(2)}$ defined on it is such that (i) $\omega^{(2)}$ is non-singular and (ii) $d\omega^{(2)} = 0$, i.e. the form $\omega^{(2)}$ is closed.

It is known that symplectic manifolds admit a special atlas in which the form $\omega^{(2)}$ (sometimes called "symplectic structure") becomes canonical: namely, for any point $P \in M^{2n}$ there exist a neighbourhood $U(P)$ and coordinates $\{p^1, \ldots, p^n, q^1, \ldots, q^n\}$ in which $\omega^{(2)}$ can be written as $\omega^{(2)} = \sum\limits_{i=1}^{n} dp^i \wedge dq^i$. These coordinates are called *symplectic*. Their existence is demonstrated by the Darboux theorem which will not be proved here, for we confine ourselves only to an example of a symplectic manifold. Let $M^{2n} = \mathbf{R}^{2n}$, let $p^1, \ldots, p^n, q^1, \ldots, q^n$ be Cartesian coordinates in $\mathbf{R}^{2n}$, and let $\omega^{(2)} = \sum\limits_{i=1}^{n} dp^i \wedge dq^i$. Since the coefficients $\omega_{ij}^{(2)}$ of this form are constant, $d\omega^{(2)} \equiv 0$, i.e. the form is closed. Its non-singularity is obvious and, therefore, $\mathbf{R}^{2n}$ with the form $\omega^{(2)}$ becomes a symplectic manifold. Another important example of a symplectic manifold is a smooth two-dimensional orientable manifold, i.e. S^2 with a certain number of handles. Such a manifold admits a symplectic structure if by the

2-form of interest we mean an exterior form of volume relative to the Riemannian metric, i.e. the form defined in the local coordinates (x, y) as $\omega^{(2)} = \sqrt{g}\, dx \wedge dy$, where $g = \det g_{ij}$. Then M^2 becomes a symplectic manifold. Indeed, $\omega^{(2)}$ is non-singular because $\sqrt{g} \neq 0$; furthermore, $\omega^{(2)}$ is, apparently, closed.

Remark. The Darboux theorem on the existence of a symplectic atlas is not a trivial consequence of the possibility of reducing a 2-form to a canonical "block" structure at each point. The Darboux theorem asserts that such a reduction can be performed in the entire neighbourhood, and this fact needs special demonstration.

Definition. A smooth vector field X on a symplectic manifold M^{2n} with the form $\omega^{(2)}$ is called *Hamiltonian* if this field is of the form $\mathbf{X} = \operatorname{sgrad} H$, where H is a smooth function on M^{2n}. The function H is called the *Hamiltonian of a flow*.

Let us find an explicit coordinate expression for the operation $\overrightarrow{\operatorname{sgrad}} f$. We have: $\omega(\overrightarrow{\operatorname{sgrad}} f, \mathbf{Y}) = \omega_{ij} (\operatorname{sgrad} f)^i Y^j = Y^k \frac{\partial f}{\partial x^k}$, whence $(\operatorname{sgrad} f)^i = \omega^{ij} \frac{\partial f}{\partial x^j}$, i.e. the vector $\operatorname{sgrad} f$ is dual (relative to the scalar product ω) to the covector $\operatorname{grad} f = \left(\frac{\partial f}{\partial x^i}\right)$. Here is an example of the Hamiltonian field. Let H be an arbitrary smooth function on M^{2n} referred to symplectic coordinates (p^i, q^i), i.e. $\omega^{(2)} = \sum_{i=1}^{n} dp^i \wedge dq^i$, where the coordinates (p^i, q^i) are Cartesian. The operation $\operatorname{sgrad} f$ is then of the form: $\operatorname{sgrad} H = \left(\frac{\partial H}{\partial q^i}, -\frac{\partial H}{\partial p^i}\right)$, i.e. the field $\mathbf{X} = \left(\frac{\partial H}{\partial q^i}, -\frac{\partial H}{\partial p^i}\right)$, $1 \leqslant i \leqslant n$, is Hamiltonian. This example clearly shows the distinction between Hamiltonian and potential fields: a Hamiltonian field is of the form $\left(\frac{\partial H}{\partial q^i}, -\frac{\partial H}{\partial p^i}\right)$ and a potential field $\operatorname{grad} H = \left(\frac{\partial H}{\partial p^i}, \frac{\partial H}{\partial q^i}\right)$. In particular, these fields (for the same function H) are orthogonal with respect to the Euclidean scalar product. Some fields, which are Hamiltonian for a function H, may be potential for some other function H.

A Hamiltonian field admits an important description in terms of a one-parameter group of diffeomorphisms generated by this field. Let $\mathbf{X} = \operatorname{sgrad} H$ be a Hamiltonian field and let $\mathfrak{G}_t$ be the corresponding group of shifts along the integral trajectories of the field. This group acts on the form $\omega^{(2)}$, sending it into another, generally distinct, 2-form $(\mathfrak{G}_t)^* \omega^{(2)}$ which can be represented as follows. If the mapping $\mathfrak{G}_t$ sends a point y to $x_t(y) = \mathfrak{G}_t(y)$, the coordinates $(x_t^i(y))$ may be considered as smooth functions of (y^j); we have

then

$$(\mathfrak{G}_t)^* \omega^{(2)} = \omega_{ij}(x_t(y))\, dx_t^i(y) \wedge dx_t^j(y)$$
$$= \omega_{ij}(x_t(y)) \frac{\partial x_t^i(y)}{\partial y^k} \frac{\partial x_t^j(y)}{\partial y^p} dy^k \wedge dy^p = \omega_{kp}(t, y)\, dy^k \wedge dy^p,$$

where $\omega_{kp}(t, y) = \omega_{ij}(x_t(y)) \frac{\partial x_t^i(y)}{\partial y^k} \frac{\partial x_t^j(y)}{\partial y^p}$ (see Fig. 7.26).

A vector field $\mathbf{X}$ on a symplectic manifold may not admit global representation in the form $\mathbf{X} = \operatorname{sgrad} H$, but it may possess the

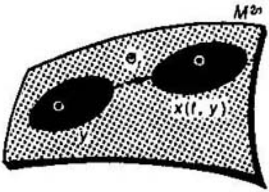

Figure 7.26

following important property: for any point $P \in M^{2n}$ there exist a neighbourhood $U(P)$ and a function H_U defined on this neighbourhood such that $\mathbf{X}(\alpha) = \operatorname{sgrad} H_U(\alpha)$, where $\alpha \in U$. The field $\mathbf{X}$ will be called locally Hamiltonian. Each neighbourhood U requires its own Hamiltonian H_U, but these Hamiltonians cannot, in general, be "matched" into a single Hamiltonian defined on the entire manifold. An analysis of locally Hamiltonian fields is thus reduced to studying Hamiltonian fields in $\mathbf{R}^{2n}$ with the structure $\sum_{i=1}^{n} dp^i \wedge dq^i$, since the neighbourhood U can be mapped diffeomorphically onto a domain in the symplectic manifold $(\mathbf{R}^{2n}, \omega^{(2)})$.

Theorem 4. *A vector field $\mathbf{X}$ on a symplectic manifold M^{2n} is locally Hamiltonian if and only if the one-parameter group $\mathfrak{G}_t$ preserves the symplectic structure $\omega^{(2)}$.*

Proof. We prove the theorem only for the case where M^{2n} is $\mathbf{R}^{2n}$ with the structure $\omega^{(2)} = \sum_{i=1}^{n} dp^i \wedge dq^i$. The general case is reduced to this one by the Darboux theorem (its proof is omitted).

We now demonstrate that if $\omega^{(2)}$ is preserved under the action of $\mathfrak{G}_t$ the field $\mathbf{X}$ is locally Hamiltonian. That $\omega^{(2)}$ is mapped into itself under the action of $\mathfrak{G}_t$ means that the derivative of the form along the field $\mathbf{X}$ vanishes ($\mathbf{X}$ is the velocity field for the action of

then

$$(\mathfrak{G}_t)^* \omega^{(2)} = \omega_{ij}(x_t(y))\, dx_t^i(y) \wedge dx_t^j(y)$$
$$= \omega_{ij}(x_t(y)) \frac{\partial x_t^i(y)}{\partial y^k} \frac{\partial x_t^j(y)}{\partial y^p}\, dy^k \wedge dy^p = \omega_{kp}(t, y)\, dy^k \wedge dy^p,$$

where $\omega_{kp}(t, y) = \omega_{ij}(x_t(y)) \frac{\partial x_t^i(y)}{\partial y^k} \frac{\partial x_t^j(y)}{\partial y^p}$ (see Fig. 7.26).

A vector field $\mathbf{X}$ on a symplectic manifold may not admit global representation in the form $\mathbf{X} = \operatorname{sgrad} H$, but it may possess the

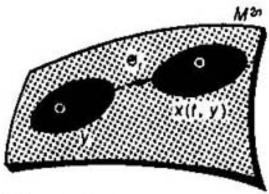

Figure 7.26

following important property: for any point $P \in M^{2n}$ there exist a neighbourhood $U(P)$ and a function H_U defined on this neighbourhood such that $\mathbf{X}(\alpha) = \operatorname{sgrad} H_U(\alpha)$, where $\alpha \in U$. The field $\mathbf{X}$ will be called locally Hamiltonian. Each neighbourhood U requires its own Hamiltonian H_U, but these Hamiltonians cannot, in general, be "matched" into a single Hamiltonian defined on the entire manifold. An analysis of locally Hamiltonian fields is thus reduced to studying Hamiltonian fields in $\mathbf{R}^{2n}$ with the structure $\sum_{i=1}^{n} dp^i \wedge dq^i$, since the neighbourhood U can be mapped diffeomorphically onto a domain in the symplectic manifold $(\mathbf{R}^{2n}, \omega^{(2)})$.

Theorem 4. *A vector field* $\mathbf{X}$ *on a symplectic manifold* M^{2n} *is locally Hamiltonian if and only if the one-parameter group* $\mathfrak{G}_t$ *preserves the symplectic structure* $\omega^{(2)}$.

Proof. We prove the theorem only for the case where M^{2n} is $\mathbf{R}^{2n}$ with the structure $\omega^{(2)} = \sum_{i=1}^{n} dp^i \wedge dq^i$. The general case is reduced to this one by the Darboux theorem (its proof is omitted).

We now demonstrate that if $\omega^{(2)}$ is preserved under the action of $\mathfrak{G}_t$ the field $\mathbf{X}$ is locally Hamiltonian. That $\omega^{(2)}$ is mapped into itself under the action of $\mathfrak{G}_t$ means that the derivative of the form along the field $\mathbf{X}$ vanishes ($\mathbf{X}$ is the velocity field for the action of

$\mathfrak{G}_t$), i.e. $0 = \frac{d}{dt}\,\omega^{(2)}(\gamma(t))$, where $\gamma(t)$ is an integral trajectory of **X**. We have

$$\begin{aligned}\frac{d}{dt}\,\omega^{(2)}(\gamma(t)) &= \frac{d}{dt}\sum_{i=1}^{n} dp^i(t)\wedge dq^i(t)\\ &= \sum_{i=1}^{n}\left[\frac{d}{dt}(dp^i(t))\wedge dq^i(t)+dp^i(t)\wedge\frac{d}{dt}dq^i(t)\right]\\ &= \sum_{i=1}^{n}\left[d\left(\frac{dp^i(t)}{dt}\right)\wedge dq^i(t)+dp^i(t)\wedge d\left(\frac{dq^i(t)}{dt}\right)\right]\\ &= \sum_{i=1}^{n}[d(X^i(t))\wedge dq^i(t)+dp^i(t)\wedge dY^i(t)],\end{aligned}$$

where $\mathbf{X}=(X^i, Y^i)$, $1\leqslant i\leqslant n$. Also,

$$\begin{aligned}&\frac{d}{dt}\,\omega^{(2)}(\gamma(t))\\ &= \sum_{i=1}^{n}\left[\left(\frac{\partial X^i}{\partial p^k}dp^k+\frac{\partial X^i}{\partial q^k}dq^k\right)\wedge dq^i\right.\\ &\qquad\left.+dp^i\wedge\left(\frac{\partial Y^i}{\partial p^k}dp^k+\frac{\partial Y^i}{\partial q^k}dq^k\right)\right]\\ &= \sum_{i=1}^{n}\left[\left(\frac{\partial X^i}{\partial p^k}+\frac{\partial Y^k}{\partial q^i}\right)dp^k\wedge dq^i\right.\\ &\qquad\left.+\left(\frac{\partial X^i}{\partial q^k}-\frac{\partial X^k}{\partial q^i}\right)dq^k\underset{(k<i)}{\wedge}dq^i+\left(\frac{\partial Y^i}{\partial p^k}-\frac{\partial Y^k}{\partial p^i}\right)dp^i\underset{(i<k)}{\wedge}dp^k\right]=0,\end{aligned}$$

i.e.

$$-\frac{\partial X^i}{\partial p^k}=\frac{\partial Y^k}{\partial q^i},\quad \frac{\partial X^k}{\partial q^i}=\frac{\partial X^i}{\partial q^k},\quad \frac{\partial Y^i}{\partial p^k}=\frac{\partial Y^k}{\partial p^i}.$$

This means that the 1-form $\omega^{(1)}=\sum_{i=1}^{n}(Y^i\,dp^i-X^i\,dq^i)$ is closed. Indeed, according to the relations derived above,

$$d\omega^{(1)} = d\sum_{i=1}^{n}(Y^i\,dp^i - X^i\,dp^i)$$

$$= \sum_{i=1}^{n}\Big[\Big(\frac{\partial Y^i}{\partial p^k}\,dp^k + \frac{\partial Y^i}{\partial q^k}\,dq^k\Big)\wedge dp^i$$

$$-\Big(\frac{\partial X^i}{\partial p^k}\,dp^k + \frac{\partial X^i}{\partial q^k}\,dq^k\Big)\wedge dq^i\Big]$$

$$= \sum_{i=1}^{n}\Big[\Big(\frac{\partial Y^i}{\partial p^k} - \frac{\partial Y^k}{\partial p^i}\Big)\,dp^k \underset{(k<i)}{\wedge} dp^i$$

$$+\Big(\frac{\partial V^i}{\partial q^k} + \frac{\partial X^k}{\partial p^i}\Big)\,dq^k\wedge dp^i + \Big(\frac{\partial X^k}{\partial q^i} - \frac{\partial X^i}{\partial q^k}\Big)\,dq^k\wedge dq^i = 0.$$

Since the 1-form $\omega^{(1)}$ is defined on $\mathbf{R}^{2n}$, it is closed and therefore there exists on $\mathbf{R}^{2n}$ a smooth function H such that its gradient (ordinary grad) is equal to this form, i.e. $\omega^{(1)} = dH$. Indeed, H can be expressed explicitly in terms of the form $\omega^{(1)}$, suffice it to set

$$H(x) = \int_{\gamma}\omega^{(1)} = \int_{\gamma(t)}\sum_{i=1}^{n}(Y^i\,dp^i - X^i\,dq^i) = \int_{0}^{1}\sum_{i=1}^{n}\Big(Y^i\,\frac{dp^i}{dt} - X^i\,\frac{dq^i}{dt}\Big)\,dt,$$

where $\gamma(t)$ is an arbitrary smooth trajectory in $\mathbf{R}^{2n}$ from a fixed point (say, the point 0) to a variable point (p^i, q^i). The number $H(x)$ does not depend on the path from the point 0 to the point $x = (p^i, q^i)$. For $n = 2$ this fact has already been proved in Chapter 4. For $n > 2$ it follows from Stokes' formula. Indeed, consider two paths $\gamma_1(t)$ and $\gamma_2(t)$, where $\gamma_1(0) = \gamma_2(0)$, $\gamma_1(1) = \gamma_2(1)$, then

$$\int_{\gamma_1}\omega^{(1)} = \int_{\gamma_2}\omega^{(1)} = \int_{\gamma_1\cup(-\gamma_2)}\omega^{(1)} = \int_{\partial D^2}\omega^{(1)} = \int_{D^2}d\omega^{(1)} = 0,$$

where D^2 is an arbitrary disk bounded by the trajectories γ_1 and γ_2 (the disk may have self-intersections). This completes the construction of H. It remains to prove the converse statement: if $\mathbf{X}$ is a Hamiltonian field in $\mathbf{R}^{2n}$, the group $\mathfrak{G}_t$ preserves $\omega^{(2)}$, i.e. $\mathbf{X}(\omega^{(2)}) = 0$. To prove this fact, it suffices to reverse all the calculations just made. The theorem is proved.

Using this result, we can easily give examples of Hamiltonian flows. For instance, let us consider a sphere with g handles provided with Riemannian metric and symplectic structure $\omega^{(2)} = \sqrt{\det g_{ij}}\,dx\wedge dy$ (see above). How can we describe all Hamiltonian flows on this symplectic manifold? We recall that a field $\mathbf{X}$ on M_g^2

has zero divergence if the area of an arbitrary domain A remains unchanged under shifts along the integral trajectories of the field.

Corollary 1. *A flow on M_g^2 is Hamiltonian if and only if it is incompressible, i.e.* div $X = 0$.

Proof. According to Theorem 4, we should describe all the flows that preserve the form $\omega^{(2)}$. Since by this form we mean the form of volume, $\omega^{(2)} = \sqrt{\det g_{ij}}\, dx \wedge dy$, the Hamiltonian flow is the field that preserves the form of volume. As we know from Chapter 4, this field represents flows of zero divergence, and only these flows. The corollary is proved.

Let us consider, for instance, a two-dimensional sphere S^2 represented by a completed complex straight line $\mathbf{R}^2 \cup \infty$. In this case, we can find on S^2 a wide variety of Hamiltonian flows (with respect to the standard metric on S^2): it suffices to consider the fields grad Re $f(z)$ and grad Im $f(z)$, where $f(z)$ is a complex-analytic function of the variable $z = x + iy$. Hence, all incompressible (and irrotational) flows considered in Chapter 4, Sec. 4.4, are Hamiltonian. The symplectic structure on S^2 referred to the coordinates (x, y) (which are used for all "finite" points) is given by $\omega^{(2)} = dx \wedge dy$. However, there also exist on S^2 Hamiltonian flows which are not irrotational (see below).

Let us consider an important example of Hamiltonian flow, the dynamical system which describes the motion of a three-dimensional rigid body with a fixed point. We should consider $\mathbf{R}^3$ identified with the linear space of all real-valued skew-symmetric matrices of order 3, suppose $X \in \mathbf{R}^3$ is such a matrix. Define in $\mathbf{R}^3$ a self-adjoint linear operator φ: $\mathbf{R}^3 \to \mathbf{R}^3$ such that $\varphi X = XI + IX$, where I is the matrix $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$ and (λ_i) are real numbers. Apparently, the matrix φX is skew-symmetric. Let us consider the system of differential equations $\varphi \dot{X} = [X, \varphi X]$ (written in matrix form). It is known from mechanics that this system describes the motion of a three-dimensional rigid body with a fixed point. Let us write the equations in coordinate form. It is convenient to consider an n-dimensional rigid body rotating in $\mathbf{R}^n$ about a fixed point. This rotation is described by the system $\varphi \dot{X} = [X, \varphi X]$, where φ: $so(n) \to so(n)$, $so(n) = (X\colon X^T = -X)$, φ is of the form $\varphi X = XI + IX$, $I = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$.

Assuming $X = (x_{ij})$, $x_{ij} = -x_{ji}$, we obtain

$$(\varphi X)_{ij} = (\lambda_i + \lambda_j)\, x_{ij},$$

$$(\dot{X})_{ij}=\sum_k (x_{ik}(\varphi X)_{kj}-(\varphi X)_{ik}x_{kj})$$
$$=\sum_k (x_{ik}(\lambda_k+\lambda_j)x_{kj}-(\lambda_i+\lambda_k)x_{ik}x_{kj}=(\lambda_j-\lambda_i)\sum_k x_{ik}x_{kj},$$
$$\dot{x}_{ij}=\frac{\lambda_j-\lambda_i}{\lambda_j+\lambda_i}\sum_k x_{ik}x_{kj}.$$

For $n=3$ we have

$$\dot{x}_{12}=\frac{\lambda_2-\lambda_1}{\lambda_2+\lambda_1}x_{13}x_{32},\quad \dot{x}_{13}=\frac{\lambda_3-\lambda_1}{\lambda_3+\lambda_1}x_{12}x_{23},\quad \dot{x}_{23}=\frac{\lambda_3-\lambda_2}{\lambda_3+\lambda_2}x_{21}x_{13}.$$

Setting $x_{12}=x$, $x_{13}=y$, $x_{23}=z$, we arrive at the following system:

$$\dot{x}=\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}yz,\quad \dot{y}=\frac{\lambda_3-\lambda_1}{\lambda_3+\lambda_1}xz,\quad \dot{z}=\frac{\lambda_2-\lambda_3}{\lambda_2+\lambda_3}yx.$$

This flow (in $\mathbf{R}^3$) admits two integrals: $P(x, y, z)=x^2+y^2+z^2$ and $Q(x, y, z)=x^2(\lambda_1+\lambda_2)+y^2(\lambda_1+\lambda_3)+z^2(\lambda_2+\lambda_3)$. Indeed, to prove that these two functions are constant along the integral trajectories of the flow, it is sufficient to calculate the derivatives $\dot{X}(P)$ and $\dot{X}(Q)$. We have

$$\dot{X}(P)=2xyz\left(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}+\frac{\lambda_3-\lambda_1}{\lambda_3+\lambda_1}+\frac{\lambda_2-\lambda_3}{\lambda_2+\lambda_3}\right)$$
$$=\frac{2xyz}{(\lambda_1+\lambda_3)(\lambda_1+\lambda_2)(\lambda_2+\lambda_3)}[(\lambda_1-\lambda_2)(\lambda_3+\lambda_1)(\lambda_2+\lambda_3)$$
$$+(\lambda_3-\lambda_1)(\lambda_1+\lambda_2)(\lambda_2+\lambda_3)+(\lambda_2-\lambda_3)(\lambda_1+\lambda_2)(\lambda_3+\lambda_1)]=0,$$
$$\dot{X}(Q)=2xyz(\lambda_1-\lambda_2+\lambda_3-\lambda_1+\lambda_2-\lambda_3)\equiv 0,$$

which is what was required. If $\lambda_1, \lambda_2, \lambda_3>0$ and λ_i are pairwise distinct, these two integrals are functionally independent (at the points of general position) (verify!). Hence, the integral trajectories are the lines of intersection of two families of surfaces, spheres $P=$ const and ellipsoids $Q=$ const. It is convenient to picture the field X on the level surface $Q=$ const (see Fig. 7.27). Fix a sphere $P=$ const and restrict the flow $\dot{X}$ to this sphere. We assert that this flow is Hamiltonian. Indeed, it is required to prove that X preserves the form of volume, i.e. that $\dot{X}$ has zero divergence. It suffices to demonstrate that the flow preserves the form of volume in the surrounding Euclidean space. But this is obvious because the expression $\frac{\partial \dot{x}}{\partial x}+\frac{\partial \dot{y}}{\partial y}+\frac{\partial \dot{z}}{\partial z}$ is identically zero (see the explicit form of $\dot{x}$, $\dot{y}$, $\dot{z}$). This example shows that integrals of a flow play a key role in the description of flow behaviour. In the case of a three-dimensional rigid body

the equations of motion are completely integrable, i.e. we have found two integrals of motion. The equations of motion are also completely integrable (in a certain exact sense) for an n-dimensional rigid body; but the proof of this fact is far from being trivial and is beyond the scope of this book.

Now a few words about the integrals of Hamiltonian flows. On the space of smooth functions on a symplectic manifold it is useful to introduce a new operation generated by the symplectic structure.

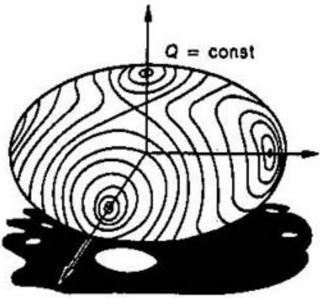

Figure 7.27

Definition. The *Poisson bracket of two smooth functions f and g on a symplectic manifold M^{2n}* is the function $\{f, g\}$ defined by

$$\{f, g\} = \omega^{(2)}(\operatorname{sgrad} f, \operatorname{sgrad} g) = \omega_{ij}(\operatorname{sgrad} f)^i (\operatorname{sgrad} g)^j.$$

The explicit formula for the operation $f, g \to \{f, g\}$ is written as

$$\{f, g\} = \omega^{k\alpha} \frac{\partial f}{\partial x^k} \frac{\partial g}{\partial x^\alpha} \text{ because } (\operatorname{sgrad} f)^i = \omega^{i\alpha} \frac{\partial f}{\partial x^\alpha}.$$

Proposition 1. *The operation $f, g \to \{f, g\}$ satisfies the following conditions:* (1) *it is bilinear,* (2) *it is skew-symmetric, i.e.* $\{f, g\} = -\{g, f\}$ *and* (3) *the Jacobi identity*

$$\{h, \{f, g\}\} + \{g, \{h, f\}\} + \{f, \{g, h\}\} = 0$$

holds true.

Proof. We prove this theorem for a particular case of $M^{2n} = \mathbf{R}^{2n}$ with a canonical symplectic structure. Conditions (1) and (2) directly follow from the definition of a symplectic structure. Let us prove (3). Denote by L_f differentiation along the field sgrad f, i.e. $L_f(h) =$ (sgrad f) h. We assert that the expression on the left-hand side of (3) is the sum of monomials, each containing the second partial

derivative of one of the functions f, g, and h. Indeed, consider $\{f, g\}$; from the definition of sgrad we obtain

$$\{f, g\} = \omega^{(2)}(\operatorname{sgrad} f, \operatorname{sgrad} g) = -(\operatorname{sgrad} g) f$$
$$= -L_g(f) = -(\operatorname{sgrad} g)^i \frac{\partial f}{\partial x^i}.$$

The components of the field $(\operatorname{sgrad} g)^i$ are partial derivatives of the function g. Really, since $\omega^{(2)} = \sum_{i=1}^{n} dp^i \wedge dq^i$, we have $(\operatorname{sgrad} g)^i = -\frac{\partial g}{\partial q^i}$ for $1 \leqslant i \leqslant n$; $(\operatorname{sgrad} g)^j = \frac{\partial g}{\partial p^j}$ for $1 \leqslant j \leqslant n$; the matrix of the form $\omega^{(2)}$ in $\mathbf{R}^{2n}(p^1, \ldots, p^n, q^1, \ldots, q^n)$ is constant and has canonical "block" structure. Hence, $\{f, g\}$ is a linear combination of monomials which are the products of first-order partial derivatives. Thus,

$$\{h, \{f, g\}\} = -\{\{f, g\} h\} = -L_{\{f, g\}}(h) = -(\operatorname{sgrad}\{f, g\})^i \frac{\partial h}{\partial x^i},$$

i.e. the expression $\{h, \{f, g\}\}$ is the sum of monomials, each including second-order partial derivatives either of function f or of function g. Thus, we have demonstrated that the left-hand side of identity (3) is the sum of monomials, each containing, as a factor, a second-order partial derivative. Let us now fix the function h and collect all the terms on the left-hand side of (3) that contain second-order partial derivatives of h

$$\{f, \{g, h\}\} - \{g, \{f, h\}\} = L_f L_g h - L_g L_f h = [L_f, L_g]\, h.$$

On the other hand, we know that the commutator of two first-order operators, L_f and L_g, is an operator of the first (but not second!) order, i.e. the expression $[L_f, L_g]\, h$ includes only first-order derivatives of h. Thus, no second-order derivative of h appears on the left-hand side of (3). Using the same reasoning, we can easily verify that the left-hand side of (3) does not contain second-order derivatives of f or g either, i.e. the left-hand side of (3) is equal to zero. The proposition is proved.

A linear space with the bilinear operation satisfying conditions (1)-(3) is called a *Lie algebra*.

Corollary 2. *The linear space $F(M^{2n})$ of all smooth functions on a symplectic manifold (M^{2n}, ω) is a Lie algebra with respect to the Poisson bracket $\{f, g\}$.*

This follows from Proposition 1. The Lie algebra $F(M)$ is infinite-dimensional. Let us consider the mapping α of the Lie algebra $F(M)$ into the Lie algebra $V(M)$ of all smooth vector fields on M defined by the formula $\alpha(f) = \operatorname{sgrad} f$.

Corollary 3. *The mapping* $\alpha: f \to \operatorname{sgrad} f$ *is a homomorphism of Lie algebras, i.e.* $\alpha\{f, g\} = [\alpha f, \alpha g]$.

Proof. It is required to prove that $\operatorname{sgrad}\{f, g\} = [\operatorname{sgrad} f, \operatorname{sgrad} g]$. From the Jacobi identity we have

$$\begin{aligned}(\operatorname{sgrad}\{f, g\})\, h &= \omega(\operatorname{sgrad}\{f, g\}, \operatorname{sgrad} h) = -\{h, \{f, g\}\}\\ &= -\{\{g, \{f, h\}\} + \{f, \{g, h\}\} = -L_g L_f h + L_f L_g h\\ &= [L_f, L_g]\, h,\end{aligned}$$

$$\operatorname{sgrad}\{f, g\} = [\operatorname{sgrad} f, \operatorname{sgrad} g],$$

which is what was required.

Since α is a homomorphism of Lie algebras, a subalgebra, denoted by $H(M)$, is the image of the Lie algebra $F(M)$ in the Lie algebra $V(M)$ of all smooth fields on M. The definition of α implies that this subalgebra is the Lie algebra of all Hamiltonian fields on M. Thus, $\alpha: F(M) \to H(M)$ is an epimorphism. But α is not a monomorphism, for it has a non-trivial kernel: this kernel is formed by locally constant functions. If we assume that M is connected, the kernel of α is one-dimensional and consists of constant functions on M. Hence, $H(M) \cong F(M)|_{R_1}$. The subalgebra $H(M) \subset V(M)$, $H(M) = H_\omega(M)$ depends on the choice of the symplectic structure ω on M. If we introduce on M another symplectic structure ω', the subalgebra $H_{\omega'}(M) \subset V(M)$ will differ from $H_\omega(M)$. It should be noted that potential fields $\operatorname{grad} f$ are distinct from Hamiltonian fields $\operatorname{sgrad} f$. Indeed, potential fields do not form a subalgebra in $V(M)$, i.e. the commutator of two potential fields need not necessarily be a potential field, potential fields on a circle being an example. Each vector field on S^1 can be defined by a smooth function $P(\varphi)$, where φ is the angular coordinate and $P(\varphi)$ is the tangent vector component. Then, the commutator of two fields $P(\varphi)$ and $Q(\varphi)$ on S^1 is the field with the component $R = PQ'_\varphi - QP'_\varphi$ (verify!). Set $P = \cos\varphi$, $Q = \sin\varphi$, then $R \equiv 1$. Thus, we have represented the field with constant component (equal to 1) as a commutator of two potential fields. A constant field is not potential, for it has a closed trajectory without singularities.

The operation of calculating the Poisson bracket plays an important role in the study of integrals of Hamiltonian flows.

Proposition 2. *Let* $\mathbf{v} = \operatorname{sgrad} H$ *be a Hamiltonian flow on* M^{2n} *and let* $f(x)$ *be a smooth function which commutes with the Hamiltonian* H, *i.e.* $\{f, H\} \equiv 0$. *Then* f *is an integral of the flow* $\mathbf{v}$, *i.e. the function* f *is constant along the integral trajectories of the field* $\mathbf{v}$. *Conversely, any integral* g *of the field* $\operatorname{sgrad} H$ *commutes with* H, *i.e.* $\{g, H\} = 0$.

Proof. It suffices to calculate the derivative of f along the field $\mathbf{v}$, i.e. $(\operatorname{sgrad} H) f$. By the definition of $\operatorname{sgrad} H$, we have $(\operatorname{sgrad} H) f = \{H, f\} \equiv 0$, which is what was required.

Since $\{H, H\} \equiv 0$, the function H is an integral of the flow sgrad H.

Proposition 3. *Let f and g be two integrals of* sgrad H, *then* $\{f, g\}$ *is also an integral of the flow* sgrad H.

The *proof* directly follows from the Jacobi identity.

Thus, by calculating the Poisson bracket of two integrals, we obtain a new integral and the latter may, however, functionally depend on the two initial integrals.

Name Index

Subject Index

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR